

Patching together a world view

Data sets encapsulating the behaviour of the Earth system are one of the greatest technological achievements of our age — and one of the most deserving of future investment.

Technology can change the way we see the world. If the artist David Hockney is to be believed, the camera obscura changed the way artists drew things, and thus how their audiences saw them. Centuries later, photographic film changed the visual arts again, as painters sought to recapture subjectivity in fresh impressionisms and expressionisms in response to the new technology. Then cinematography brought with it a new mastery over time. Compressed, it turned buds to blooms in seconds — reversed, it re-erected falling chimneys with pleasing symbolic power. These tricks became embedded in our minds, letting us think of time moving backwards and forwards, faster and slower with an educated ease previously absent from the imagination.

In the past two decades, the computer has changed things yet again, introducing an almost infinite capacity to bring what was previously non-visual to the eye, and an almost infinite range of points of view impossible to reach in any other way. The ability to change point-of-view and depth-of-field massively and arbitrarily has created a peculiarly contemporary way of seeing, which American technology writer Steven Johnson has called “the long zoom”. This is when a camera focused on, say, a human eye appears to hurtle pell-mell through the pupil to the nucleus of a cell — or pulls back from the orbit of the eye to an orbit round the planet.

In the world of the long zoom, the planetary scale has a particular significance. It links every image of the world to the great image of Earth that contains them all. It builds on and subverts the change first introduced by space flight almost 50 years ago — the ability to stand outside what was previously seen only from within. The long zoom integrates the inside and the outside, giving computers a means of marshalling vast data sets — as users of Google Earth can testify. Geospatial imagery becomes a great uniter of data; whether the data come from satellites looking down, or sensors deep in the oceans, or tracking systems strapped to walruses or gas monitors sitting above forest canopies, computers can, in principle, put them all together (see articles starting on page 778). This is why seemingly arcane developments such as the European Union’s INSPIRE directive, a measure that tidies up access to geodata and provides an Internet portal for accessing them, are important. They set the standards by which the world can be freely reassembled.

The long view

The creation of these new ways of seeing the world would be a significant aesthetic achievement even if they had no commercial, scientific or strategic use. In fact they have all three — as well as an even greater environmental usefulness. After the expansion of human population, intensive agriculture and industrial development that marked the twentieth century, it is only with the help of global monitoring systems that today’s arrangements of everything from urbanization to epidemiology can be properly understood.

One of the most profound contributions to this approach came from the late David Keeling, a pioneer of climate research who was the first person reliably to measure carbon dioxide levels at remote locations such as Mauna Kea in Hawaii or the South Pole, in what his friend and boss Roger Revelle famously called mankind’s “great experiment” with the planet’s climate. Keeling’s simple instruments became the basis of a network around the world for monitoring trace gases. At various times it was suggested to Keeling that he should perhaps desist from taking such endless care over a single data stream — that this wasn’t the basis of great science. It took courage and conviction to keep going — and even now, his heirs struggle to continue the work in the face of unwilling funders and apathetic peers (see page 789).

Now or never

Monitoring the Earth system requires great expertise, not just to build the instruments but to use them properly and interpret their output. Many scientists are, however, far from enthused by projects that do not involve the forming and testing of hypotheses. At worst, monitoring is traduced as stamp collecting and looked down on as drudgery.

Such attitudes must not be allowed to prevail. Testing hypotheses about how the world works requires not just information on the current state of the three-dimensional globe, but on its progress through the fourth dimension of time. Data on the colour of the seas that are not gathered today can never be gathered in the future — gaps left in the record cannot be filled (see page 782). And continuous data sets are going to be vital to the validation of the ever more informative models of the Earth system that we need.

This is why operational systems for data collection in which scientists play key roles are so important. Only they can give us multiscale and multifactor ways of seeing the world that are up to the challenges of the twenty-first century. When the expenditure needed to maintain these data flows conflicts with the funds needed to support fresh scientific research, researchers must acknowledge that there is a strong case for preferring continuous, operational monitoring. An accurate and reliable record of what is going on can trump any particular strategy for trying to understand it.

There is only one Earth, with only one history, and we get only one chance to record it. Ideas not followed through can be taken up again later. A record not made is gone for good. Long zooms in and out of our ever more detailed images of Earth will delight and inform us for years to come. But no digital trickery can replace the steady, fateful pan from past to future. ■

“Ideas not followed through can be taken up again later. A record not made is gone for good.”

Africa's academies

Robust scientific institutions won't be built in a day.

Last month in Dakar, Senegal, representatives of scientific academies, science ministers, donors and development specialists gathered to take stock of a multi-year project to develop stronger scientific academies in Africa, led by the US National Academies and funded by the Bill & Melinda Gates Foundation.

The African Science Academy Development Initiative (ASADI) has been undertaken on the assumption that both scientists and nations stand to benefit from the existence of robust, well respected and reasonably independent institutions that can speak forcefully and effectively for science. Elsewhere in the world, such bodies serve to sustain excellence within the scientific community itself, to foster informed public discourse on science-related issues and to provide policy-makers with sound advice on these issues, encouraging rational decision-making.

Some solid progress has been made since ASADI got under way in 2004. Initially, the initiative has concentrated on providing intensive support for academies in South Africa, Nigeria and Uganda. These projects are each at different stages of development. South Africa already had an academy structure, albeit one fragmented by the legacy of apartheid. Nigeria has a troubled history and vast, unmet needs. Uganda has a young academy, striving to make its presence felt in a relatively open political environment. ASADI also provides more modest support to academies in Senegal, Ghana, Cameroon and Kenya, and to the African Academy of Sciences in Nairobi.

The goals of the project are more abstract than those of most development efforts in science and medicine. But the establishment of credible institutions (such as scientific academies) is needed to underpin broader economic and social development and the dearth of such institutions is a critical problem in much of Africa.

Organizations such as the Royal Society in London and the National Academies in Washington DC have taken many decades to build up their respective positions. Although occasionally lampooned as self-important associations of elderly white males, they are in fact important components of good governance. Their voices command respect, even in the most disrespectful corners of the media, and their advice is widely sought.

ASADI is providing long-term support to establish staff and premises for its three target academies, which are starting to do useful work. The Academy of Science of South Africa, for example, produced a comprehensive study in August on AIDS and tuberculosis. The Ugandan academy is reporting on the use of DDT to control mosquitoes indoors. The Nigerian academy has held workshops on blood safety and infant mortality. Getting governments to follow through on such work remains a major challenge, however.

The African academies need to establish a level of recognition with their respective publics and governments, and then to make that recognition stick in a climate where influence is often ephemeral. Building up that recognition will require careful, unspectacular legwork over many years. The Gates Foundation will consider continuing funding for the project next year and will do well — if the quality of the emerging academies' work is high enough — to support it over the long haul. ■

Shared genomes

Nature's publishers are expanding the licence for reuse of genome papers.

Although *Nature* and the Nature journals are built on a business model funded by subscribers and other sources of revenue, various initiatives have been implemented to enhance the accessibility of the research papers published in these journals.

They have long been freely available to researchers in the 100 or so poorest countries through the World Health Organization's Hinari initiative and others like it. Machine access is being enhanced by the open text-mining initiative of the Nature Publishing Group (NPG) (<http://opentextmining.org>). Preprints of original versions of papers can be deposited in arXiv (<http://arxiv.org>) and *Nature Precedings* (<http://precedings.nature.com>) without compromising their acceptability for publication. And final authors' versions of papers can be deposited in PubMed Central and other public servers from six months after publication. Authors retain copyright of their work, whereas NPG retains the licence to publish it.

For many years, a more generous arrangement has been made for papers reporting full genome sequences. (The paper reporting the sequence and analysis of 12 species of *Drosophila* is the most recent example, see *Nature* 450, 203; 2007). These papers are freely

accessible on NPG's website from the moment of publication. This recognizes a consistent character of 'genome' papers: they represent the completion of a key and fundamental research resource, describing and reflecting on what has been revealed but not usually providing insights into mechanism. Although some papers in other disciplines might also be characterized in this way, the fundamental character of the genome has led NPG to make a systematic exception.

In the continuing drive to make papers as accessible as possible, NPG is now introducing a 'creative commons' licence for the reuse of such genome papers. The licence (see www.nature.com/authors/editorial_policies/license.html) allows non-commercial publishers, however they might be defined, to reuse the pdf and html versions of the paper. In particular, users are free to copy, distribute, transmit and adapt the contribution, provided this is for non-commercial purposes, subject to the same or similar licence conditions and due attribution.

In 1996, as human genome sequencing was getting under way, leading players stated: "It was agreed that all human genomic sequence information, generated by centres funded for large-scale human sequencing, should be freely available and in the public domain in order to encourage research and development and to maximise its benefit to society" (see www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml). These principles have continued to guide the field, and NPG has consistently made genome papers freely available in keeping with them. This new licence allows us to formalize the arrangement. ■

RESEARCH HIGHLIGHTS

Regal revolution

Proc. R. Soc. B doi:10.1098/rspb.2007.1422 (2007)

Cape bees (*Apis mellifera capensis*) are the only species of honeybee in which workers can produce new queens. Cape bee workers do this by means of a weird reproductive trick that allows them to lay eggs that develop into queens even though the eggs are not fertilized. Madeleine Beekman, of the University of Sydney in Australia, and her team have figured out how often these workers compete with a resident queen to produce future monarchs.

In most bee societies, the queen lays all the eggs in the hive. In a few, workers can lay eggs that develop into males. Because an average worker is genetically more similar to the queen's sons than to the sons of other workers, it will remove the male offspring of its own caste. But the daughter queens of Cape bee workers remain. Other workers, in this case, are as related to those daughter queens as they are to the queens born of the current sovereign.

The authors' genetic analyses show that 23 of 39 new queens from seven colonies in South Africa were the offspring of workers, and that 15 of these 23 were produced by parasitic workers, from other Cape bee colonies. Furthermore, three new queens born of resident queens were produced asexually, suggesting that queens 'choose' how they reproduce.



B. OLDROYD

MATERIALS SCIENCE

Diamond insight

Science **318**, 1424–1430 (2007)

Researchers say that they have resolved the puzzle of why the surface of diamond (pictured below), usually an outstanding insulator, can conduct electricity when exposed to air.

The phenomenon has perplexed chemists ever since it was first observed in 1989. Various groups have shown that the material's conductivity is due to 'holes' of positive charge, but the holes' origin has been harder to establish. One controversial idea is that they result from the movement of electrons from carbon atoms into a thin film of water adsorbed onto the diamond's surface.

By carefully measuring the acidity and oxygen concentrations of diamonds in aqueous solution, John Angus of Case Western Reserve University in

Cleveland and his colleagues show that electron transfer between diamond and water does occur. This transfer both creates the conducting holes and assists water adhesion to the surface of the normally hydrophobic diamond.

PLANT BIOLOGY

Burly burley

Proc. Natl Acad. Sci. USA

doi:10.1073/pnas.0709453104 (2007)

During periods of low rainfall, plants often age their leaves and shed them, in part to avoid losing water through leaf stomata. By inserting a gene that suppresses leaf ageing into the genome of *Nicotiana tabacum*, researchers have engineered a tobacco plant that can tolerate extreme drought.

Eduardo Blumwald of the University of California, Davis, and his colleagues added a gene encoding isopentenyltransferase, an enzyme that catalyses the rate-limiting step of a senescence-inhibiting pathway, and linked it to a genetic switch that turns 'on' when plants are drought-stressed. Their transgenic tobacco recovered quickly after 15 days without water, whereas wild-type tobacco died. Moreover, cutting the water supply of modified plants by 70% of the amount normally required reduced yields only slightly.

These findings could benefit tobacco growers with rain-fed farms in countries such as Malawi, where burley tobacco is an important cash crop and climate-change models predict less and more inconsistent precipitation.

ASTROPHYSICS

Star stoppers

Mon. Not. R. Astron. Soc.

doi:10.1111/j.1365-2966.2007.12487.x (2007)

Something must limit how large galaxies grow. In galaxies that have one, the activity of a supermassive black hole seems a likely candidate, but collecting supporting data for this theory has proved tricky. Kevin Schawinski at the University of Oxford, UK, and his colleagues have made the first measurements linking black holes to the end of star birth in galaxies.

The team studied almost 16,000 galaxies from the Sloan Digital Sky Survey database, analysing the spectral features of the light emitted by each galaxy's stars. They found that galaxies with brighter disks of gas around their black holes — that is, those with more active black holes — contain fewer new stars than do those with dimmer disks, or less active black holes.

But how black holes might halt galaxy growth remains a mystery. One idea is that the disks heat galactic gas, preventing stars from coalescing. Another points to jets of gas along the polar axes of the disks disrupting star coalescence.

MOLECULAR BIOLOGY

Two-way control

Science doi: 10.1126/science.1149460 (2007)

MicroRNAs are well-characterized as molecules that can silence genes — often hundreds at once — by selectively preventing the expression of messenger RNAs by



E. NATHAN/ALAMY

ribosomes or by targeting them for destruction. A few examples of microRNAs inducing gene expression have been reported, but the paucity of details as to how they might do so has prompted scepticism.

Working with human embryonic kidney cells, Joan Steitz and her colleagues at Yale University School of Medicine in New Haven, Connecticut, have determined that a microRNA called miR369-3 brings three proteins together, including one with an important role in intracellular communication. This turns up protein expression.

Steitz and her team then looked at two often-studied microRNAs and found that they also stimulate protein production on cell-cycle arrest, leading them to propose that this may be a common function of microRNAs. Their work adds to evidence linking microRNAs to various cancers.

COSMOLOGY

Listening to inflation

Phys. Rev. Lett. **99**, 221301 (2007)

The very early Universe is widely believed to have expanded rapidly during a process known as inflation. Richard Easther and his co-workers at Yale University in New Haven, Connecticut, calculated that the end of this inflationary period would have excited gravitational waves that we may soon have the technology to detect.

They model how, immediately after inflation, the Universe would have oscillated like a ball coming to rest at the bottom of a bowl. Its oscillations would have produced background gravity waves 'loud' enough to be picked up by planned observatories, such as future iterations of the Big Bang Observer and the Laser Interferometer Gravitational

Wave Observatory. According to the team's calculations, these instruments should provide a rare glimpse of inflation itself, and might help to discriminate between different theoretical models of how it happened.

GEOLOGY

Fresh advances

Geology **35**, 1075–1078 (2007)

Periods of massive runoff from melting land ice have been linked to rapid shifts in global climate, through changes in the circulation of the northern Atlantic Ocean. New research led by Norway's Geological Survey suggests that these freshwater incursions might be more common than previously thought.

Jochen Knies, of the Geological Survey, and his colleagues analysed the oxygen and carbon isotopes in a seabed sediment core drilled in 1993, at a point off Greenland's east coast where the Arctic and Atlantic oceans meet. The researchers expected to find evidence of freshwater peaks during glacial to interglacial transitions. But instead they uncovered freshwater discharges spread liberally over the past 800,000 years, including some during periods when the global climate was relatively stable.

These events could be helpful to climate modellers seeking to shed light on abrupt climate changes, the researchers suggest.

ZOOLOGY

Advantageous offspring

Biol. Lett. doi:10.1098/rsbl.2007.0482 (2007); *Proc. R. Soc. B* doi:10.1098/rspb.2007.1401 (2007)
Evolutionary theory posits that sexual creatures maximize their number of descendents if they have more sons than daughters when conditions are optimal,

and more daughters than sons in less ideal circumstances. Yet the cues that mammals use to achieve this are unclear.

Samuli Helle, of the University of Turku in Finland, and his colleagues report a correlation between the annual mean temperature in northern Finland and the sex ratio of newborns in three populations of indigenous Sami people (pictured below) for the years 1745–1890. The team compared demographic data from the parish registers of Lutheran churches with a climatic record reconstructed from tree rings and an index of the North Atlantic Oscillation. Warmer years brought an increased proportion of boys, in keeping with theory.



Meanwhile, Elissa Cameron, of the University of Pretoria in South Africa, and her colleagues raised the proportion of daughters in mouse litters by lowering the blood glucose levels of females during conception. The researchers added dexamethasone, a steroid that blocks glucose uptake into the blood, to the drinking water of female mice for three days while the animals had access to a mate. Only 41.9% of the litters of dexamethasone-treated mice were male, compared with 53.5% in control litters.

JOURNAL CLUB

Sarah E. Hitchcock-DeGregori, Robert Wood Johnson Medical School, Piscataway, New Jersey

A molecular biophysicist muses on how a vital structural protein might have turned out differently.

One thing I love about science is its surprises. One signature eukaryotic protein is actin, a cytoskeletal protein involved in cell migration and muscle contraction. Few would have guessed that actin has a similar structure to proteins with little sequence homology or shared function beyond hydrolysis of the

energy-storage molecule ATP.

Some prokaryotic cytoskeletal proteins share actin's folding pattern and, like actin, can form filaments. Electron microscopy initially indicated that the filament structure of the DNA-segregation protein ParM, for example, is similar to actin's. With improved methods, Orlova *et al.* revealed another surprise: ParM's helix winds in the opposite direction to actin's (A. Orlova *et al. Nature Struct. Mol. Biol.* **14**, 921–926; 2007). The main differences between actin and ParM filaments are in contact regions between subunits, which are crucial for constructing this new

molecular model of the filament.

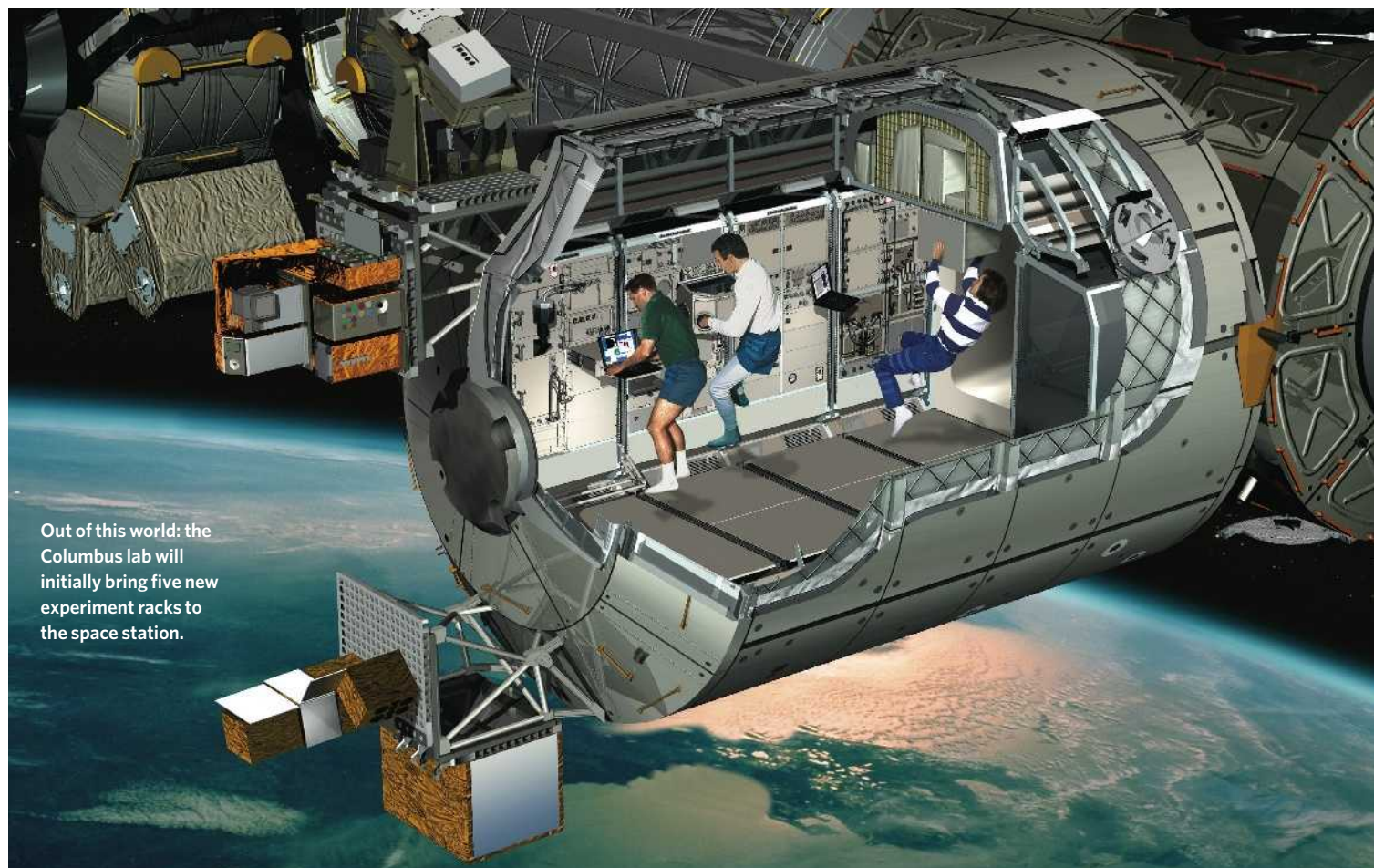
If we consider each evolutionary change as a 'mutagenesis experiment', the number required for a common ancestor to become actin or ParM is mind-boggling. There would be new functions and failures along the way — even a single amino-acid substitution in haemoglobin can cause sickle-cell anaemia. Without a record of genetic changes we cannot know the intermediate successes and failures, as we do for some bizarre invertebrates found in the fossil record. Stephen Jay Gould recounts these discoveries in *Wonderful Life*, in which he warns of being bound

by conventional thinking.

Eukaryotic life settled on the actin filament, constrained by evolution to be dynamic, to work with its motor protein, myosin, and myriad binding proteins. Just think, if actin filaments had different inter-subunit contacts — such as those in ParM — myosin would be unable to bind to it, an important helical protein, tropomyosin, would coil in the wrong direction for binding along the filament and ... well, it's like me wondering what I would be like if I had a different father.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

NEWS



Out of this world: the Columbus lab will initially bring five new experiment racks to the space station.

Columbus begins voyage of discovery

After years of delay, Europe's largest contribution to the International Space Station (ISS) is finally poised to head into orbit.

If all goes to plan, the space shuttle Atlantis will lift off with the Columbus research module on 6 December from the Kennedy Space Center in Cape Canaveral, Florida. Over an 11-day mission, American and European astronauts will connect the module to the ever-growing ISS.

Costing €700 million (US\$1 billion), Columbus is the single largest part of the station to be provided by the European Space Agency (ESA). Within its 75-cubic-metre shell is room for ten payload racks to hold experiments. ESA has already installed five racks for work in a range of fields, including biology, fluid science and physiology. The module also has mounting points for four external payloads, two of which will be fitted during the mission: one to test materials in the vacuum of space and another to observe the Sun.

The launch ends a lengthy delay for Colum-

bus, which a decade ago was scheduled to enter orbit at the start of the new millennium. The station's construction slipped by several years before it came to a stop in 2003, when the break up of the space shuttle Columbia during re-entry halted all shuttle flights (see *Nature* **421**, 561–562; 2003). Assembly only resumed two years later once NASA had completed a full safety review of the shuttle fleet.

The setbacks have left ESA's research hopes partially grounded. The agency has an 8.3% stake in the US\$70-billion–100-billion station, but without facilities of its own, European experiments had to be conducted in either American or Russian modules. The complicated barter system under which the station runs meant that using those segments came at a price in man-hours and other resources, according to Marc Heppener, ESA's chief scientist for the station. "Having our

own real-estate really makes things nicer."

After President George W. Bush announced plans in 2004 to send astronauts to the Moon, NASA redirected its science on the ISS to applied problems of space travel. ESA, meanwhile, has maintained a broader set of fundamental research questions. Among the first experiments to run in Columbus will be a simulation of geophysical fluid flow under microgravity. The experiment will use two spheres, one inside the other, with a fluid

between them, and an electrical field to simulate gravity at the centre. Astronauts will spin the spheres to create fluid-flow patterns of the sort that take place in Earth's atmosphere and oceans.

But many of the larger and more complex experiments planned will have to be deferred, according to Martin Zell who is head of research at ESA's human space-flight research directorate.

"Now that the infrastructure is there, we'd better use it in the best possible way."

D. DUCROS/ESA

**HAVE YOUR SAY**

Comment on any of our news stories, online.

www.nature.com/news

The curtailment of the shuttle has hindered the ability to transfer crew, leaving the ISS staffed with just three of the anticipated seven astronauts. Because maintaining the station is time-consuming, those astronauts can devote to research only about a quarter of the time originally allotted for science. The European portion of that time is smaller still — with Columbus in place, astronauts will have between 15 and 20 hours every six months to devote to European experiments, Zell says. That means some time-intensive studies of biology and human physiology will have to wait until 2009, when the station's crew is scheduled to double to six.

Just a year later, Columbus's scientific schedule will face another setback: NASA's retirement of the space shuttle. Without the shuttle, there will be no good way to return samples to Earth, says Julie Robinson, NASA's ISS programme scientist at the Johnson Space Center in Houston, Texas. Russia's Soyuz capsule, which will temporarily replace the shuttle, can hold just 50 kilograms of extra weight on its way home. Biological samples in particular, which can require bulky refrigeration, are likely to mount up at the station unless another return method can be developed, she says. "It's going to be a huge problem."

Heppener says that ESA is now "working hard" on technologies that will allow astronauts to do more of the required analyses at the station. Robinson adds that new commercially developed vehicles might allow for more samples to be sent home.

The setbacks will only increase the long-standing scepticism of space-station science held by some researchers. Studies of human physiology in space are useful only if governments are willing to pursue costly programmes of exploration beyond Earth orbit, says Paul Murdin at the Institute of Astronomy in Cambridge, UK. Other microgravity experiments are only of "specialized interest", he adds. "The money being spent could be better used for something else."

But others maintain that there are some fundamentally interesting experiments that could be done. For example, a weightless environment could help physicists to study Bose-Einstein condensates, clumps of ultracold atoms that act identically. It could also be used to improve atomic-laser experiments and the accuracy of atomic clocks, according to Günther Hasinger of the Max Planck Institute for Extraterrestrial Physics in Garching, Germany. "Now that the infrastructure is there," he says, "we'd better use it in the best possible way."

Geoff Brumfiel

UN talks only one part of the negotiations needed on climate

As delegations from around the world converge this week on Bali, Indonesia, for the latest round of United Nations (UN) climate talks, one thing is increasingly clear: although the UN Framework Convention on Climate Change is indisputably the main venue for such talks, it is by no means the only one. Nor should it be, many experts say.

For instance, last week French President Nicolas Sarkozy led a delegation visit to China at which key topics included energy and global warming, and the French nuclear giant Areva walked away with a deal for two nuclear power plants and the possibility of many more in the world's fastest-growing economy.

And a week before that, more than a dozen Asian nations, including India and China, signed an agreement to push for clean energy and tackle global warming. There are also coalitions aimed at urging the international community to include tropical-forest protection in whatever treaty succeeds the Kyoto Protocol. And global warming has been a primary focus of at least three international meetings this year — those of the G8 industrialized countries, the UN and, for the first time, a meeting under the auspices of US President George W. Bush in Washington DC.

"I'm actually not so worried about the treaty negotiations. I think it's much more important to get a small number of countries around a table and work out a game plan in each individual case," says David Victor, who heads Stanford University's Program on Energy and Sustainable Development.

The goal heading into Bali is not to negotiate the treaty itself but to settle on a roadmap for negotiations, with a likely end date of 2009. From this perspective, there simply won't be enough time to work out all of the details at UN meetings. In fact, once the roadmap is set, much of the work on any proposals made could be formulated well in advance of the meetings themselves.

"There's pretty broad agreement that the Framework Convention is the place to cut the grand deal, but that will only happen once the major players reach a stronger consensus," says Elliot Diringer, director of international strategies at the Pew Center on Global Climate Change in Arlington, Virginia. "Bilateral contact and other initiatives outside the UN process are key to reaching that consensus," he adds, as long as everyone keeps their eye on achieving a global treaty.

Many advocacy groups bristled when the Bush administration held its meeting of "major emitters" in September, fearing that it was intended to undercut the UN negotiations.

Rob Stavins, an environmental economist at Harvard University, acknowledges there is room for scepticism towards Bush, who has dragged his feet in the international negotiations for years. But bringing nations such as China, India and Brazil to the table with the industrial nations makes sense, he says. "It's questionable whether there was much that was meaningful in that meeting, but it was the right set of parties around the table."

The developing nations know that their bargaining power stems from fears among Western nations that their own actions to curb greenhouse gases will mean little unless everyone participates. There is also growing scepticism about the "clean development mechanism", the Kyoto Protocol's main vehicle for carbon-reducing technology transfer to developing



Presidential duo: China's Hu Jintao (left) and France's Nicolas Sarkozy met to discuss climate change.

nations. Victor says crafting alternative incentives for nations such as China might prove intractable without a forum that is "smaller and a lot more flexible" than the UN process.

"My guess is that the Chinese are going to be a whole lot more comfortable in bilateral and multilateral negotiations than under the klieg lights of the UN process," he says. "Those kinds of discussions aren't antithetical to Kyoto — they are just more important. Big treaties almost always follow such negotiations rather than lead them."

John Ashton, a UK climate-change envoy, says debate about venue and process misses the point. Global warming should top every agenda as nations meet in the coming years, but the only venue for resolving the issue is the UN. "An international treaty is an expression of political will," he says. "It's not about the precise architecture of the treaty. There are lots of available policies, and we understand them quite well — it's the urgency and amplitude with which we apply those policies that's important."

Jeff Tollefson

E. FEFERBERG/POOL/AP

SPECIAL REPORT

Solar power: California's latest gold rush

Green-energy companies are enjoying a boom in investment. But will they live up to expectation, asks **Declan Butler**.

Silicon Valley is greening. Investors are flocking to low-carbon (clean) energy technologies, fuelling a boom in the sector, with investments set to overtake those in Internet start-ups. But does this venture-capital explosion herald another dotcom bubble?

Last week, Google announced its entrance into the green-energy field — hundreds of millions of dollars for a technology push to make renewable energies cheaper than coal. And in the same month, Nobel prizewinner Al Gore and his London-based firm Generation Investment Management joined forces with one of Silicon Valley's foremost green-energy investors, Kleiner Perkins Caufield & Byers in Menlo Park, California — the company that nurtured the likes of Amazon, Google and Genentech — to create an alliance to fund global climate solutions.

For the fast-moving entrepreneurs of the valley, who have successfully invested in disruptive technologies such as the Internet to change the face of entire industries, the next frontier is the roughly US\$6-trillion energy market, where the dinosaurs of power-generation utilities have traditionally invested a pittance in research and development. "Venture capital is exactly what we need to try new things outside the bounds of what the traditional energy companies think is worth doing," says Vinod Khosla, a veteran entrepreneur who co-founded Sun Microsystems and now heads Khosla Ventures in Menlo Park, one of the most prominent clean-energy venture-capital firms. "There is almost no technology risk-taking in any of the energy companies." Khosla predicts that within five years there will be a green form of electricity that is cheaper than coal, and cleaner fuels that are cheaper than oil.

The US venture-capital industry has spent \$2.6 billion on clean-energy technologies in the first three-quarters of this year — up from \$1.8 billion last year, and \$533 million in 2005 — according to new figures released on

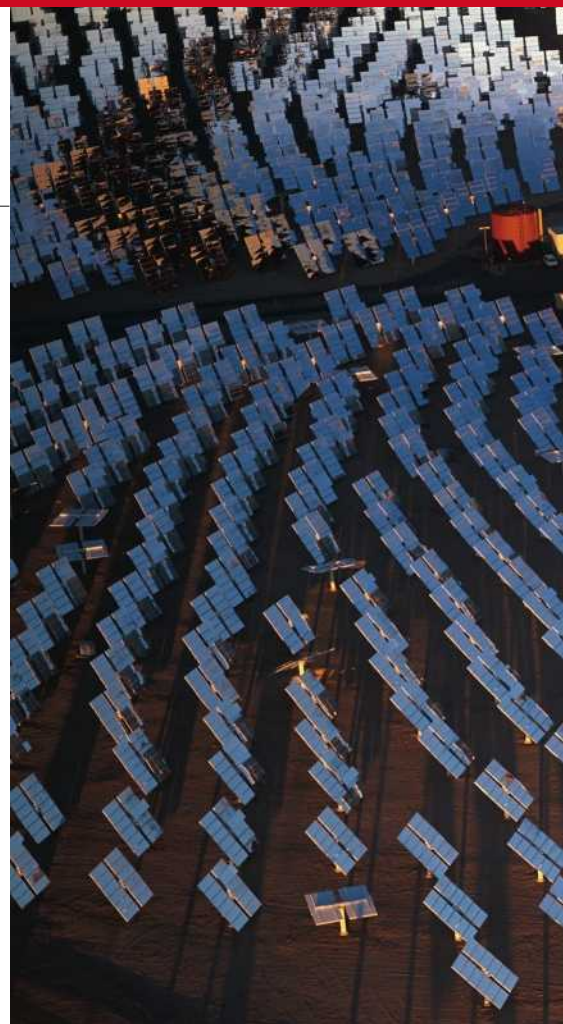
28 November by Thomson Financial and the US National Venture Capital Association (NVCA), which represents almost 500 venture-capital funds. Venture capital is a thermometer of a new sector, dealing as it does with investment in new technologies, products and services. And for every dollar invested here, many more are spent in operations, building such things as solar generators, wind farms and bio-fuel plants. Total worldwide investment in clean tech, including such investments, jumped from \$28.3 billion in 2004 to \$75.4 billion last year. This year, that figure has already climbed to \$94.5 billion, says Philip Verey at the London-based consultancy firm New Energy Finance.

"Don't make the mistake of looking for the future in your rear-view mirror."

— Vinod Khosla

Worldwide, the star attraction for venture-capital investors is solar power. Although other forms of renewable energy can make significant contributions to current markets, only sunlight is available in the amounts required to substitute completely for the energy quantities currently derived from hydrocarbons. This year's US figures show the same trend, with solar attracting the largest share of investments, \$664.6 million, followed by biofuels and smart power systems (see *Nature* **445**, 586–588; 2007).

Although the United States is lagging far behind European countries such as Denmark and Germany in implementing renewables, its venture-capital investments in clean tech now more than double those in Europe. California scooped \$726.2 million of this year's US clean-tech venture funding, followed by



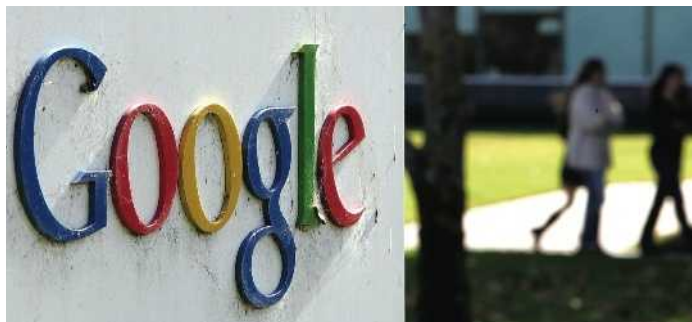
Massachusetts (\$292.6 million) and Texas (\$149.4 million). Almost \$1 billion of US investment went abroad, including a \$200-million investment in Brazil's Brazilian Renewable Energy, which produces ethanol, and a \$118-million investment in China's Yingli Green Energy Holding Company, which makes photovoltaic solar systems.

On the up

The increases significantly buck a trend — total public and private spending on energy research in the United States and elsewhere has been dropping steadily since the 1970s. In an analysis of energy-research spending published in January (G. F. Nemet and D. M. Kammen *Energy Policy* **35**, 746–755; 2007), the authors estimate that the US invests \$1 billion less in energy R&D

than it did a decade ago, and that this now represents just 2% of all federal R&D, compared with 10% in the 1980s. By contrast, spending on defence and health has been increasing by 10–15% annually during that period.

The private sector's share of the shrinking energy-research pie has also dwindled, and now makes up just a quarter of investments, compared with a half in the 1980s. The surge in new investment in energy by venture capitalists, established



Google is committing millions to solar, wind and geothermal technologies.



G. STEINMETZ/CORBIS

Californian investments in solar power, such as that produced at Dagget in the Mojave Desert, are leading the market in renewable energy worldwide.

energy firms such as General Electric and new entrants such as Google is therefore “extremely important”, says Gregory Nemet of the University of Wisconsin-Madison. “It’s difficult to envision successfully addressing concerns about energy independence and climate change without fully engaging the capabilities, resources and human ingenuity that these entities can apply,” he says.

But he cautions that the growth in green investment needs to be seen in perspective. For example, the entire US venture-capital investment in clean tech in 2006 — at \$1.8 billion — was exactly the same as that spent on R&D by the biotech company Genentech. “The magnitudes of the challenges of energy independence and climate change are so large that we are still orders of magnitude away from devoting the societal resources we need to deal with them,” Nemet says.

There is always a chance that the current wave of investment could peter out, perhaps owing to a substantial fall in oil and energy prices, or a fading of environmental concerns — but these are unlikely. The biggest risk is that the pace of basic technological improvements may fail to provide a pipeline of emerging technologies that

venture capitalists can feed off. Venture capitalists are not in the business of funding the basic research that will be needed to make the sort of breakthroughs needed to make solar energy cheaper than coal. Without a significant expansion of public spending on basic energy research, the innovation pipeline risks drying up.

The clean-tech market is “fraught with pitfalls and not for the inexperienced or the faint of heart”, according to NVCA president Mark Heesen. It is fit only for investors ready to look long-term and with a deep knowledge of the sector, he warns. “Short-term ‘tourists’ should steer clear.”

But for the moment, investor interest shows no sign of waning, and a string of new companies is preparing initial public offerings. The WilderHill New Energy Global Innovation Index (NEX), which tracks the share performance of new energy companies, is outperforming indices such as the P&P 500 and NASDAQ. Clean energy is already creating its billionaires: Shi Zhengrong, for example, who in 2001 created the company Suntech, making solar cells, is now China’s second wealthiest man.

The investments being made will allow development of many new technologies, with

rapid natural selection leaving fewer survivors. “Ninety per cent of the companies that exist today in this segment won’t be around in five years,” predicts Martin Roscheisen, chief executive of Nanosolar in San José, California, which tests new materials in solar-cell design (see *Nature* 443, 19–22; 2006). But this period of exuberance will be vital to spawning the next generation of low-carbon energy giants, he argues.

And if the bubble bursts? “Bubbles happen when over-exuberant investors who don’t understand the business get in, and I am certain one will happen in energy too,” says Khosla. “But just as in the dotcom world, good companies will continue to be built before and after the bubble.”

Venture capitalists will no doubt create stunningly successful new energy companies — the Googles and Amazons of tomorrow’s lower-carbon world — but only time will tell whether these can be sufficiently disruptive to the gargantuan energy industry on the massive scale needed to affect climate change. Solar energy and other renewables hardly make it above the x-axis on a bar chart of world-energy use, after all. But, Khosla advises, “don’t make the mistake of looking for the future in your rear-view mirror”.

Declan Butler

“Ninety per cent of the companies that exist today in this segment won’t be around in five years.”
— Martin Roscheisen



HOW TO MAKE A ZOMBIE COCKROACH

Researchers discover how wasps' venom makes roaches their slaves.

www.nature.com/news

LIBER SAT
LAGUNA DESIGN/SPL

Model predicts structure of crystals

"One of the continuing scandals in the physical sciences is that it remains impossible to predict the structure of even the simplest crystalline solids from their chemical composition." So wrote the chemist and former *Nature* editor John Maddox in 1988 (see *Nature* 335, 201), who was bemoaning the lack of a computational method to predict how molecules will arrange themselves in the solid state.

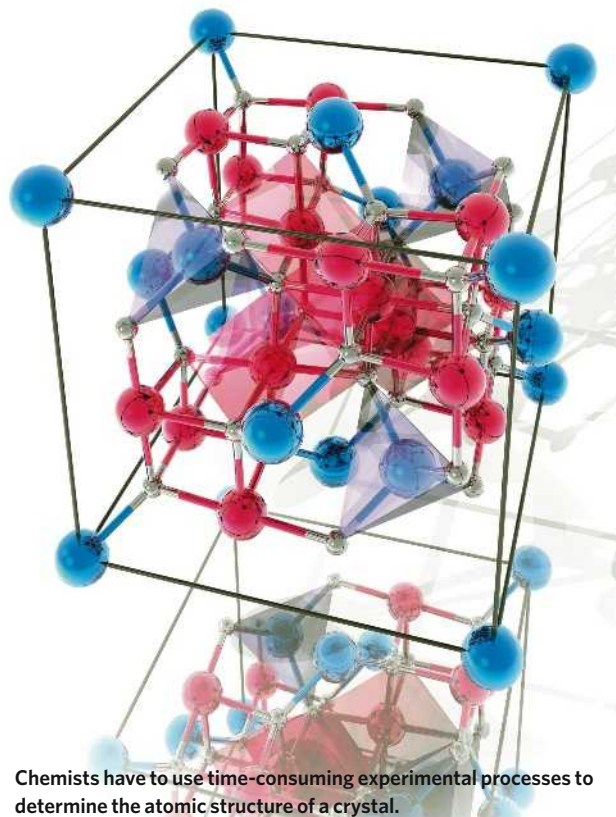
At present, chemists use X-ray crystallography to determine how atoms are arranged in a molecule and how molecules pack into a crystal. It is a time-consuming method that has remained practically unchanged for almost a century, and it means that experimentalists need to produce a high-quality crystal. Software that could predict the three-dimensional structure of compounds could allow the properties of materials, or potential drug candidates, to be examined and the different possible molecular configurations to be screened in advance, before they are made experimentally.

The problem has been attacked head-on by the Cambridge Crystallographic Data Centre (CCDC) in Britain. Every three years since 1999, the centre has set a challenge for software developers to predict the structure of four molecules, the structure of which was known only to the CCDC. But nobody had been able to predict the correct structures of all four molecules.

This year saw a breakthrough. A team consisting of Frank Leusen and John Kendrick of the University of Bradford, UK, and Marcus Neumann of Avant-garde Materials Simulation in Saint-Germain-en-Laye, France, correctly predicted the structure of all four molecules. "We have made a big step forward," says Leusen.

The 15 teams performing this year's test were each given a sketch of four molecules that showed all the atoms present and some details about how they were crystallized. These were all unpublished small organic molecules, of between 8 and 33 atoms. From this starting point, the teams ran calculations about the possible positions of the atoms and molecules within a crystal.

In most branches of chemistry, it is not enough to know a molecule's formula. For



Chemists have to use time-consuming experimental processes to determine the atomic structure of a crystal.

instance, to predict a compound's solubility, colour or how it will be taken up by the body, chemists need to know how the molecules align in a crystal. Even the smallest compound may have hundreds of thousands of possible arrangements of its molecules in a solid crystal. The structure depends on the energies of the atoms both within and between the molecules; and also on the relative positions of the atoms and molecules. The structure that scientists try to predict is the one with the lowest value for these energies: the most stable configuration. "Modelling the growth of a crystal is a daunting thing," says the CCDC's Graeme Day. "The main problem is when the molecule has options for its own shape as well as for how it packs."

"The main problem is when the molecule has options for its own shape as well as for how it packs."

Finding the most stable structure is especially important in the pharmaceutical industry. If one crystal arrangement, or polymorph, is overlooked, and it happens to be more soluble than the others, then a patient could end up taking a greater dose of a drug than is needed. And because patents are granted for a specific polymorph, companies risk being gazumped

if they don't know the entire range possible.

The winning predictive method used a new approach. "We took a route that is different from everyone else," says Neumann, who wrote the program that the team used. The team added a quantum mechanical round of calculations in between two sets of the more usual classical simulations. The first molecular-mechanics step screens the possible energies of the crystals and ranks them in order of stability. The new step then calculates a force field of the bond energies for the most stable configurations, which helps to whittle down the original list of possible structure candidates to the most likely 100 or so. "Now we can do this additional refinement that re-ranks the structures," says Leusen. Finally, this re-ranked list goes through another round of molecular-mechanical calculations based on the lattice energy and stability of the molecule.

Flexible functions

And the system worked, even for the trickiest test. Day had thrown into the test a co-crystal of two molecules, as well as a compound that was long and floppy. "In the first three tests we had no successes for flexible compounds," says Day. But in this year's test, three teams got it right.

"This is very welcome news," says David Baker, at the University of Washington in Seattle, who works on predicting the structure of proteins. But he cautions that the technique isn't necessarily applicable to larger molecules such as proteins because of the computing power needed to perform quantum calculations on large systems. Even very small proteins have hundreds of atoms — insulin, for instance, has nearly 800. And the largest proteins can have hundreds of thousands of atoms.

Neumann says that at the moment their technique will work for only 5% of the molecules that are useful for the pharmaceutical industry. The next goal is use the structure to accurately predict a crystal's properties, he says.

Maddox greeted the achievement with enthusiasm. The problem of structure prediction "stuck out like a sore thumb" 20 years ago, he says. But he isn't surprised that it has taken so long to crack. "Science doesn't necessarily move at the speed of jets."

Katharine Sanderson

ON THE RECORD

“The crocodile is showing no signs of life. We are not specialists and, to be honest, we don’t know whether it’s dead or alive.”

Rudimentary herpetology proves to be a bit too much for Nikolai Ranga, the Ukrainian emergency official charged with looking after Godzilla, recently recaptured after six months on the run from a travelling zoo.

ZOO NEWS

Royal tigers

Rather like Europe’s royal families, tigers at a breeding centre in China are having their family tree drawn up to ensure their bloodline remains pure. Keepers are DNA-testing the centre’s 800 tigers to guard against ‘intermarriage’ that could dilute the tigers’ precious gene pool.

SCORECARD



Battling obesity

US health officials say obesity levels may be hitting a plateau — although they’re still running at one in three for both sexes.



Battling incompetence

Rhode Island Hospital has been fined \$50,000 after the third instance this year of brain surgeons operating on the wrong side of a patient’s head.

NUMBER CRUNCH

50 is the number of endangered humpback whales scheduled to be caught by Japanese researchers this year — a quota that has angered conservationists.

11,000 is the number of potential names suggested for the migrating whales being tracked as part of Greenpeace International’s campaign to save them. Supporters can vote online for their favourites.

75% is the proportion of voters who chose ‘Mister Splashy Pants’ as their preferred name.

Sources: Reuters, Xinhua.net, AP, Greenpeace International

Saudi-Italian biomedical institute gets go-ahead

Saudi Arabia has come a step closer to attaining a world-class level of research with the first signed agreement between the kingdom and a Western biomedical institute.

On 26 November, two oncology institutes in Milan, Italy, signed a Memorandum of Understanding with representatives of the Saudi Arabia General Investment Authority (SAGIA). The institutes will help to train Saudi Arabian students, advise Saudi authorities on the creation of a hospital and cancer research centre and collaborate in research projects.

Saudi Arabia, which has not invested much in basic science in the past, is now actively trying to buy its way into the world of cutting-edge research. Earlier this year, it founded the King Abdullah University of Science and Technology (KAUST) with a US\$10-billion endowment (see *Nature* 447, 758–759). It also has plans to create from scratch four new ‘economic cities’. The planned research hospital will be part of the King Abdullah Economic City being developed on a 168 square-kilometre site close to the Red Sea near the cities of Jiddah, Mecca and Medina.

The SAGIA wants to model its research activities on Milan’s IFOM-IEO campus,

“Many more women are working in Saudi Arabian hospitals and research centres.”

which was created in April when the FIRC Institute of Molecular Oncology (IFOM) and the European Institute of Oncology (IEO) moved to one site. The campus is the core of a consortium there that includes university institutes and biotech firms, and incorporates research training, technology development and technology transfer. Marco Foiani, a scientific director at the IFOM, says that SAGIA got things moving very quickly after first making contact with their campus in spring this year.

Italian scientists visited Saudi Arabia in summer and will start teaching at universities there next year. “We will take the opportunity to scout for talent for our PhD programmes in molecular medicine, nanomedicine and bioethics,” Foiani says.

The two sides have agreed that Saudi students should be recruited for the IFOM-IEO’s highly competitive, international PhD programmes from next October. Saudi Arabia has little experience in biomedicine, but senior scientists at the Milan campus insist that they will not lower standards to favour one nation. “It’s a difficult part of the story — but we are committed to finding a way to implement it,”

Asia plans first cancer network

Cancer researchers from around Asia met in Nanjing, China, last month to hammer out plans for a regional network to coordinate epidemiology data and prevention.

The network would gather data from cancer registries in countries from the Philippines to Turkey — an area that has two-thirds of the world’s population and more than half of its 7.6 million cancer deaths each year, according to the World Health Organization. Most of these nations have registries, but the data are not always accurate or standardized, says Kazuo Tajima, director of the Aichi Cancer Centre Research Institute in Nagoya, Japan, who is one of the meeting’s organizers. “There is currently no way to compare notes,” he says.

The Asian Cancer Registry and Information Network, as it is being mooted, would establish the region’s first hub — probably in Japan — to hold standardized data, which could then be

used for epidemiological research, cancer risk assessment and prevention planning. It would offer some of the less-developed countries in the region a chance to catch up with modern diagnostic techniques and offer comparisons that might help figure out why, for example, incidence of liver cancer is so high in Mongolia.

The network would also offer an opportunity to tease out the role of Asian genes in the



Cancer education can break down cultural barriers.



**AMERICAN SOCIETY
FOR CELL BIOLOGY**
Find meeting reports in
our conference blog.
<http://blogs.nature.com/news>



The King Abdullah Economic City will be home to a world-class research facility.

says Pier Giuseppe Pelicci, head of experimental oncology at the IEO.

Another difficulty, says Foiani, will be the problems that women researchers face in a country that allows them so little freedom of movement. "Culturally, it is another world," he says. But Manar Al-Moneef, head of SAGIA's life-sciences sector, says that she expects a "big improvement" in conditions for women soon. "Many more women are working in hospitals and research centres than a few years ago — things are getting better rapidly," she says.

Saudi and Italian scientists will meet next month to sort out the details, and to identify the

research programmes on which they will collaborate. One such programme could be breast cancer, says Foiani. The disease has a high priority in Saudi Arabia, where intermarriage within families has led to an above-average rate of breast cancers with unusual genetic mutations.

Financial arrangements for the initiative have not yet been negotiated. A few months ago SAGIA signed a \$5.3-billion deal with international partners, including Fraunhofer research institutes in Germany to develop Bio-Sphere, a biotechnology science park planned for the Abdullah city.

Alison Abbott

development of some cancers and the reaction to particular drugs. "Most of the data used in cancer studies are from Westerners," says Sumio Sugano, a genomics specialist at the University of Tokyo who attended the International Union against Cancer (UICC) symposium in Nanjing. "This is a chance to use Asian data."

The organizers anticipate difficulties stemming from concerns over the privacy of data. "We have to move one step at a time," says Sugano. "Building trust among researchers is the first step." Organizers of the meeting hope that the network's activities can bridge some of the animosity present in the historically fractious region. The meeting was funded by Japan's Ministry of Health, Labour and Welfare, partly to celebrate the 35th anniversary of normalization of Sino-Japanese diplomatic relations. The next meeting will be in Manila in March.

Norie Kawahara, a research fellow at the Research Center for Advanced Science and Technology at the University of Tokyo, says the site location, Nanjing, where hundreds of

thousands of Chinese were massacred by the Japanese military 70 years ago — an event that is denied by some right-wing historians in Japan — was a symbolic choice.

A few years ago, when Kawahara started pushing the idea of a Japan-led Asian network, she says people refused, thinking that it sounded like a return to the Greater East Asian Co-Prosperity Sphere, the euphemistic name that Japan gave to its colonizing activities in the region. She hopes that through such activities as the anti-smoking, anti-cancer campaign she ran in a parallel session at last month's meeting, targeted at Chinese and Japanese children, she can "use science to rewrite history".

Tajima is applying for ¥60 million (US\$542,000) over 3 years to push Japan's role as a hub for the network in the future. "But we won't need gigantic sums of money," says Malcolm Moore, head of the UICC Asia regional office in Bangkok. "What is needed is for people to get together and participate."

David Cyranoski

Fund boost for German science

Science was the big winner in Germany's budget, approved last week. The Christian Democrat–Social Democrat majority approved a record 9.8% increase in the federal budget for science and education for 2008.

The German coalition government had promised to boost investment in science and, almost uniquely for a budget proposal, this passed through parliament without any cuts. The Ministry for Education and Research (BMBF) will receive €9.35 billion (US\$13.7 billion) next year — a whopping €150 million more than the government had initially proposed in July.

Germany's research council, the DFG, which funds most university research, will see a 3% rise next year in its €1.6-billion budget. Non-university research organizations, including the Max Planck Society, the Fraunhofer Society, which runs centres for applied research, and the Helmholtz Association, which runs Germany's 15 national research centres, will also

receive around 3% more money each. The BMBF also directly funds basic and applied science in fields of strategic importance, including health and climate research. Its grants will increase by 18% to €3.5 billion, including €285 million for winning universities in the government's 'excellence initiative' (see *Nature* 450, 452–453; 2007).

"This is a substantial growth," says Matthias Kleiner, the president of the DFG. "It's good to see that awareness of the importance of science is on the rise in Germany."

But although financial conditions for science have improved, scientific salaries in Germany are not internationally competitive, warns Rudi Balling, scientific director of the Helmholtz Centre of Infection Research in Braunschweig. "There's fresh money available to do many exciting things here, but we run the risk of losing the best people just because they can earn so much more abroad."

Quirin Schiermeier

"It's good to see that awareness of the importance of science is on the rise."

Safer way to make human stem-like cells revealed

Shinya Yamanaka of Kyoto University in Japan has refined his method for making human skin cells behave like embryonic stem cells so that it does not use a cancer-causing agent.

Late last month, Yamanaka reported making pluripotent cells — cells that can turn into any of the roughly 220 cell types in the body — by using retroviruses to carry four genes into human skin cells (K. Takahashi *et al. Cell* 131, 861–872; 2007). The four genes reprogrammed the cells into a state similar to that of cells in the early embryo. But one of the genes used to make the cell lines is *c-myc*, which can cause cancer.

Yamanaka has now shown that he can make these ‘induced pluripotent stem cells’ in both humans and mice by using just three factors — not including *c-myc* (M. Nakagawa *et al. Nature Biotechnol.* doi:10.1038/nbt1374; 2007).

Risk analysis for Boston lab slammed as ‘shoddy’

A US government safety assessment supporting the location of an infectious-disease lab in Boston was “not sound and credible”, says a report issued last week by the US National Research Council.

Construction of the \$178-million Boston University National Emerging Infectious Diseases Laboratory is due to finish in 2008, but one state and one federal lawsuit are challenging its opening. The facility will house research on deadly pathogens such as the Ebola and monkeypox viruses.

The report’s criticisms are aimed at a document, released by the US National Institutes of Health (NIH) in July, that evaluated the facility’s potential threat to its neighbours. It finds that the worst-case

scenarios proposed by the NIH, including an Ebola outbreak caused by an infected lab worker, were not relevant to assessing the true risk. Pathogens with more potential to spread, such as influenza, should have been chosen, the report says. It adds that the NIH assessment suffers from shoddy risk analysis and modelling. The NIH says it will consider the report.

AIDS worker’s misconduct ban belatedly comes to light

Scott Brodie, a former researcher at the University of Washington in Seattle, allegedly committed misconduct when doing AIDS research there in the late 1990s and early 2000s, and was banned from working for the university, according to *The Seattle Times*.

The news emerges only now because the newspaper had to win a court case to get a copy of the 2003 report of the investigation. Brodie, under the name John Doe, reportedly sued the university and the newspaper to prevent release of the report.

Brodie left the university in summer 2003 and is currently employed by drug company Schering-Plough, based in Kenilworth, New Jersey. A spokesman for the company says that it has just learned of the investigation.

According to one of the university investigators, geneticist Denny Liggitt, the report was handed to the National Institutes of Health’s Office of Research Integrity.

European satellite system is back on the launch pad

European transport ministers have revived plans for the Galileo satellite-navigation project. The decision comes almost six months after the collapse of a public-private consortium to build the €3.4-billion (US\$5-billion) network of 30 satellites.

Under the new agreement, the European



Europe’s GIOVE-B is being readied for launch.

Union will provide an additional €2.4 billion in unused agricultural funds to finance the system. The decision came despite opposition from Spain, which sought a larger stake in the venture.

Now that funding is confirmed, it will be up to the member states to decide how contracts will be awarded for the project, says Kai-Uwe Schrogl, secretary-general of the European Space Policy Institute in Vienna. “This is now a new game,” he says.

So far, only a single test satellite in the Galileo system, GIOVE-A, has been launched. A second, GIOVE-B, has faced repeated delays and is expected to launch in spring next year.

Californian stem-cell directors to be investigated

California officials may investigate potential conflicts of interest in the state’s \$3-billion stem-cell initiative.

State Controller John Chiang has asked for the Fair Political Practices Committee to look into accusations against Robert Klein, the chairman, and John Reed, a board member, of the California Institute for Regenerative Medicine (CIRM) in San Francisco. A public advocacy group has also called for both to resign after it emerged that Reed, on Klein’s advice, asked the CIRM to reconsider a grant decision involving a researcher at the Burnham Institute in La Jolla.

Reed is president of the Burnham Institute and intervened after the CIRM decided to rescind a previously awarded \$638,000 grant to a researcher who was not an on-site, full-time employee of the Burnham Institute.

Klein says he now realizes that Reed should never have written the letter, and that he looks forward to the investigation. “We welcome guidance so that we can learn,” he says.

Clarification

The News story ‘Loopholes in oversight rules revealed’ (*Nature* 450, 590; 2007) referred to a ruling issued by US Department of Energy attorneys in response to a Freedom of Information Act request. We would like to make it clear that the documents that were the subject of that request have not been released.

Endangered species cling on as rulings are overturned

The US government has revised seven decisions made by a former official who worked on the Endangered Species Act. Julie MacDonald resigned from the Fish and Wildlife Service in May after it emerged that she pressured agency scientists, and solicited advice on decisions from a friend in the online game *World of Warcraft*.

Preble’s meadow jumping mouse (*Zapus hudsonius preblei*) will not now lose its listing, but will be considered threatened only in Colorado (it is also found in Wyoming). Twelve endangered species of Hawaiian picture-wing fly will be able to stretch out in ‘critical habitats’ larger than the 0.4 hectares per species designated for them by MacDonald. And decisions affecting several other species, including the white-tailed prairie dog (pictured) and the arroyo toad, have been deemed flawed. But research needed to revise some of these decisions will be carried out only if funding can be found.



BUSINESS

Smart networking

Theoreticians have combined their expertise to form a network to help other scientists design materials or understand biological pathways. **Katharine Sanderson** reports.

For the past three months Joerg Schaefer, a physicist at the University of Würzburg in Germany, has been waiting for a group of theoretical spectroscopists to calculate the exact position of gold and platinum atoms in nanowires made from single atoms of the metals. When laid out on a semiconductor surface, the wires could potentially be used to transfer data at high speeds in microprocessors by acting as switches. Once the theoreticians have worked out what combination of metal and semiconductor should produce a wire that can be changed at will from insulator to conductor, Schaefer will try to make that combination in the lab, atom by atom.

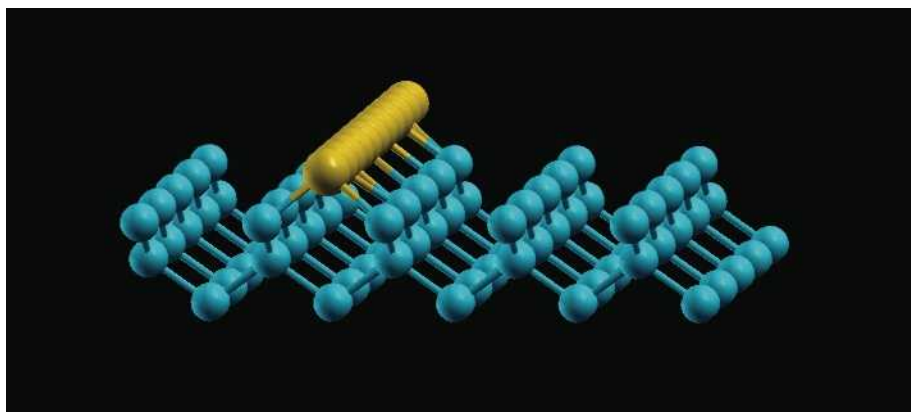
Theoretical spectroscopy simulates how electrons in molecules behave and can predict what the molecule's spectra will look like. Schaefer can't do the calculations himself, he says, because he doesn't have the necessary tools or knowledge. But he hopes that the theoretical spectroscopists will back up what he is seeing in his lab. "If everything goes right," he says, "the real world and their elusive phantom world will come to the same result."

Schaefer is one of the first clients of the European Theoretical Spectroscopy Facility (ETSF), a network created earlier this year. The network can investigate anything from how the structure of a protein affects its properties to the likely properties of a new material. Established mainly as a service for academic collaboration, its sponsors hope that it will evolve into a service for sale to industry.

Designer tools

The network simulates observational tools such as nuclear magnetic resonance, which shows how certain nuclei react to a magnetic field; infrared or Raman spectroscopy, which look at the vibration of molecules; and photoemission spectroscopy, which measures the distribution of electrons in substances. Its supporters say that it will help to save time by predicting what properties new materials or molecules will have before experimentalists try to make them — a time-consuming process.

The facility grew out of Nanoquanta, a similar network funded by the European Commission's Sixth Framework Programme, but



Spectroscopists want to know if gold nanowires (yellow) on silicon surfaces will speed up data transfer.

focused more on building its own research collaborations than on offering a service to others. Nanoquanta will run to the end of 2008 and is coordinated by Rex Godby from the University of York, UK. Both networks tackle "the theory and simulation of all sorts of characterization techniques that involve electrons," says Godby.

The ETSF doesn't have a physical home, but it involves about 120 scientists spread across 10 research institutions that are hubs for different spectroscopic techniques. To tap into the network, potential users submit a proposal that outlines their problem.

The first call for proposals, earlier this year, attracted 54 submissions from researchers in academia and industry, of which the ETSF says it will back 21.

Industrial proposals made up 15% of the submissions. One of these was from Antonio Porro, head of the Centre for Nanomaterials Applications in Construction, a firm run by Labein, a contract research and development company in Bilbao, Spain. Porro is looking at the fine structure of cements, and uses simulations to speed up the process of designing lighter, stronger materials. "We can cut by half the process of developing a material experimentally," he says.

The industrial bent of the network is mirrored by a shift in its funding as it evolves from Nanoquanta to the ETSF. Nanoquanta's funding comes from the European Commission's research directorate, which is more interested

in blue-skies research. The ETSF's main initial funding, however, is being provided by the more industry-linked information society and media directorate, to the tune of €3.7 million (US\$5.4 million).

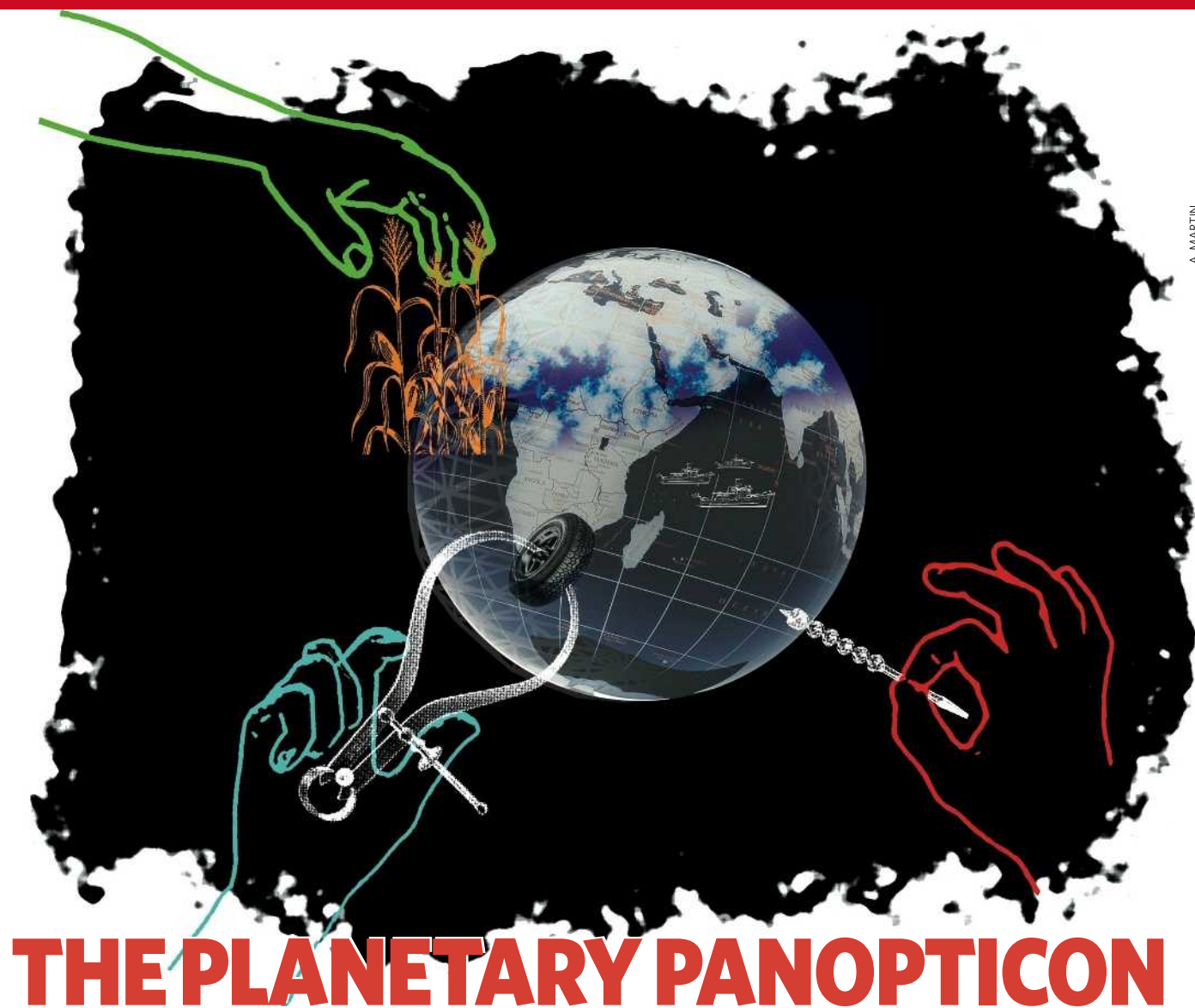
Service update

For now, the network remains very much an academic pursuit. "The aim is not to make money, it is to provide a service," says Lucia Reining, a physicist at the École Polytechnique in Paris who is the ETSF's chairwoman. But that could change. From now on the ETSF will ask for proposals every six months, and during the next couple of rounds it wants to gauge the interest from industry. "We are anxious to discover how much companies are willing to pay," says Godby.

Godby says that the ETSF has no immediate plans to spawn a start-up company: "It would be wrong for the evolution of the ETSF to be driven by our commercial users," says Godby. But a spokesman for the information society and media directorate-general says that any of the ETSF's projects that went on to produce a spin-off company would be regarded as a great success.

And Porro sees a genuine business opportunity for the venture. The ETSF "can have a specific commercial aspect," he says. "Normally, the materials design process is fully experimental," Porro says. Being able to predict and validate properties early on is a key aspect of the process, he says, and rigorous theoretical models should help with that. "This way we can save a huge amount of time and money."

M. VERSTRAETE



A. MARTIN

THE PLANETARY PANOPTICON

Technology will soon allow the world to be mapped in near-real time and at high resolution.

Declan Butler investigates the potential for operational monitoring of the planet.

Forecasting is a tricky business. You can be let down by your initial data or your model of the processes, by an unrecognized bias or just bad luck. But a dramatic forecast still has the power to grab the attention. Take this one, about the state of Earth monitoring in a couple of decades: "A user will be able to get, on demand, climate, or any other information for any place on the planet, on the land, in the oceans, or in the atmosphere, at any time, past, present and future."

The speaker is Rick Anthes, president of the University Corporation for Atmospheric Research in Boulder, Colorado, and chair of a US National Academies panel that in January released an influential 428-page blueprint on the future of Earth monitoring. The forecast he is making is based on clear and established trends: satellites are getting more cost-effective in their capabilities, and the computers and supercomputers that make use of their data are speeding up exponentially.

If that increase in technological capability

can be turned into usable systems, then the ability to monitor Earth's environment will be revolutionized. Real-time and near-real-time data will be available on soil moisture, greenhouse-gas concentrations, biological productivity, aerosol concentrations and so on, all around the world. With those data, scientists will be able to build and study models of Earth as a system far beyond what they have today.

To make that real, though, will require coherent and sustained political and institutional support, and on this front the news from the National Academies is less compelling. The US Earth-monitoring programmes are adrift without leadership, warns the academies report; the number of observational satellites and instruments has already peaked, and is set to decline over the next two decades (see page 782). No repairs are likely before the next administration, even if then. Anthes and other US scientists are keenly aware that to get a glimpse of the sort of sensible and forward-looking Earth observation strategy the

academies panel proposed, they have to look to the European Union (EU).

Europe's approach to Earth monitoring is not flashy. Its underlying philosophy flows not so much from cutting-edge research as from what amounts to weather forecasting writ large, building ever more capacity for monitoring the planet onto the day-to-day activities of meteorology — delivering data, images and products to users 24 hours a day, 365 days a year. The idea is to roll out similar easily used maps, models and forecasts on an ever-increasing range of data and processes — for example flood risks, soil and coastal erosion, crop and fish resources, air pollution, greenhouse gases, iceberg distribution and snow cover. The systems that do so would, like today's meteorological systems, generate continuous, cross-calibrated, long-term data sets on the state of the planet and its atmosphere.

It is by embedding scientific Earth-observation needs within an operational system that meets the needs of customers that the financial

case can be made for the sort of sensors attuned to various climatic and other parameters now seen only on research satellites — sophisticated spectrometers, sounders, lidars and radars. The operation of these sensors entails the recurrent costs of running fleets of satellites and sensors for decades, with regularly scheduled replacements.

A key part of the European process is the Global Monitoring for Environment and Security (GMES) programme run by the European Commission and the European Space Agency (ESA). GMES is explicitly charged with bringing the sorts of data that have previously been the province of research satellites to the citizens of Europe and beyond.

Sentinel senses

The GMES suite of 'Sentinel' satellites will be operated around the clock to routinely supply data similar to those now provided by research satellites into the foreseeable future. Sentinel 1, slated for launch at the end of 2011, will be designed for radar and build on some of the databases from Envisat, an 8-tonne ESA research behemoth that has tested out a wider range of instruments; the Sentinel 2 series will be imaging satellites with fine spectral resolution, building on the SPOT satellites; the Sentinel 3 satellites will carry forward the ocean-observing aspects of Envisat; Sentinels 4 and 5 would monitor atmospheric chemistry. The data provided by these assets would be integrated with data from future research satellites, as well as with national and international data from airborne, ground and ocean sensor webs (see *Nature* **440**, 402–405; 2006). The idea is that by the mid-2020s, Europe would have monitoring systems akin to those now in place for meteorology for all areas of environmental monitoring, says Josef Aschbacher, head of the GMES space office in Frascati, Italy, with forecasts and data on everything from global climate change to town-by-town air pollution levels.

The programme is loosely

modelled on that of the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), which supplies weather data to national met offices, and other government and commercial users. EUMETSAT is not a large organization — its annual budget is normally in the €300-million range (about US\$440 million) — but unlike ESA, or for that matter the commission, it has a track record of running operational systems in a way that works for users. The first of its Meteosat weather satellites was slotted into geosynchronous orbit in November 1977. It also now runs a weather satellite in a low-Earth orbit that complements US satellites in similar orbits. EUMETSAT recently agreed to join GMES, and is discussing directly operating future GMES satellites and ground operations, as well as hosting GMES instruments on its own weather satellites.

But GMES is a much more ambitious undertaking. It has €1.97 billion in approved funding from the EU and ESA to carry it through until 2013, covering the launching of the first three Sentinels (the contract for the first of which was signed earlier this year). In 2008 ESA will ask its member states (which differ slightly from both those of the EU, and of EUMETSAT) for a further €700 million–900 million to cover development of the Sentinel series, operations and spare spacecraft, and the commission will ask the EU for the €2.5 billion needed to operate the system until 2023.

"Climate prediction is probably the most computationally challenging problem in science."

— Tim Palmer

These requests for funding are not the only reasons that 2008 will be the crunch year for GMES, says Paul Counet, head of EUMETSAT's Strategy and International Relations Division. Many governance issues remain to be resolved,

such as who is responsible for which aspects of operations or services. And then there is the vast task of integrating national observing systems into GMES, and of finding ways for private industry to manage or add value to specific data sets and provide new services. The resolution of

these governance issues, and the forging of the long-term relationships needed to underpin operational systems, will determine whether GMES blossoms into a full-blown system or goes belly up, says Counet.

Regardless of its implementation, the operational logic of the GMES programme gets the thumbs up from across the Atlantic. "The Europeans certainly have a robust programme and are moving to make it happen," says Scott Goetz, a researcher at Woods Hole Research Center, Falmouth, Massachusetts, who uses remote sensing to model ecosystems. "GMES is a step in the right direction," agrees Kevin Trenberth, a climate researcher at the National Center for Atmospheric Research in Boulder, Colorado, who says that he would like to see a similar approach expanded in the United



Beyond the bird's-eye view: the EUMETSAT headquarters.

EUMETSAT

States and internationally, to generate a global 'climate-information system'. The fact that operational satellites must meet needs other than those of scientists — a cornerstone of the GMES approach — is no obstacle to research, he says. Researchers just need to get involved to make sure their needs are taken into account.

The data that such information systems could make available will have implications both for how well scientists can run environmental models and for what those models can do. Operational systems launched in the late-2010s and onwards are likely to generate order-of-magnitude improvements in both temporal and spatial resolution. Today, a typical weather model might have 30–50-kilometre horizontal resolution. Climate research models are even coarser — often 100–200 km. But by 2025, improvements in both data and computing will mean that weather will be modelled at 1-km resolution, and climate models at 5–10 km, predicts Anthes.

Anthes has high hopes for the impacts such

improvements could allow. "We have seen beyond a doubt in weather prediction that as the resolution increases, something almost magical starts to happen in the models, even without an increase in the observations," says Anthes. "As we go for example from a 30-km resolution model to a 5-km model, hurricane prediction, precipitation patterns and so forth become far more realistic."

More data obviously demand more computing power to make sense of them — and computing power is already the overriding limitation on how realistic models are, say many researchers. "It's a major issue," says Trenberth. To get good results you need to run the same model again and again with slightly different inputs, which eats up computing power. And every twofold increase in resolution requires a tenfold increase in the



"A user will be able to get climate information for any place on the planet at any time."

— Rick Anthes

teraflops required. "Climate prediction is probably the most computationally challenging problem in science," says Tim Palmer, a scientist at the European Centre for Medium-Range Weather Forecasts (ECMWF) in Reading, UK.

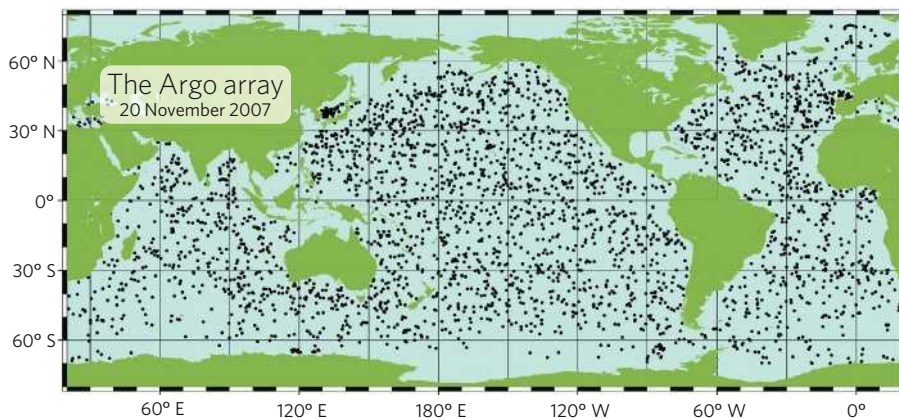
But Palmer is optimistic that the bottleneck will soon be alleviated. He points to the arrival next year of the first petaflop computers, running at peak speeds of up to 3,000 teraflops (see *Nature* 448, 6–7; 2007). By way of comparison, the ECMWF's fastest machines today run at less than a hundredth of that. Palmer predicts that by 2010, 10-petaflop machines will allow climate scientists who can get hold of them to run century-long simulations of the climate at 10-km resolution. That could be 1 km within a decade after that. Accurate modelling of cloud processes at the 1-km level, a key component currently missing from

UCAR

Observing the ocean from within

As names go, it's cleverer than most. At the mundane level, Argo stands for Array for Real-time Geostrophic Oceanography. But the name comes with not just one but two classical allusions attached. There's the *Argo*, the ship in which Jason sailed to find the golden fleece, a nod to the fact that the buoys that make up Argo were conceived as a counterpart to a satellite called Jason (see 'The Jason project'), which measures the surface topography of the oceans. And then there's Argos the giant, who was blessed with 100 eyes to see in all directions, with only a handful asleep at any one time. The modern Argo puts even that spectacularly distributed sensory system to shame — it has 3,000 different sense organs spread all around the globe (see map).

The array's 3,000-odd autonomous floats, which look like upended torpedoes, are equipped with sensors for recording temperature and salinity in the upper 2,000 metres of the ocean. Each float sinks, drifts, bobs up and transmits data to satellites on a regular basis. At its current size the array provides more than 100,000 temperature and salinity profiles each year, regardless of the season or weather. This is 20 times greater than the comparable annual measurements by research vessels and merchant ships, which in the past have been the main data source for this type of oceanography. In the past five years, Argo has



WWW.ARGONET

more than doubled the total database on some seas that ships steer clear of — such as the seas around Antarctica in winter.

Over the next years, the array's design will be reassessed with an eye to the need for additional floats and sensors. Already, 60 of the floats carry oxygen sensors — will that number be increased? What about sensors for particulate carbon, or for chlorophyll? Such extras are attractive, but always come at the cost of increased demand on the floats' batteries and thus a shorter lifetime.

Another idea is to design floats that can dive to greater depths. And special floats for polar latitudes — tethered to ice floes or equipped with acoustic sensors that tell the instrument

where it can safely surface — are already being tested in seasonally ice-covered regions.

Plugging the gaps

Predicting and closing the gaps that will constantly appear in the array will be a major logistical challenge. The types of floats currently in use have a lifetime of 4 years, meaning that around 800 instruments need to be replaced each year. And replacements need to maintain the system's global coverage. That means dropping floats out of low-flying aircraft or chartering ships to optimal sites for replenishment.

The array's annual costs — around US\$24 million — are being shared by more than 30 nations.

global climate models, could vastly improve predictions of regional climate change. Trenberth is one of a group of researchers planning to propose the creation of one or more international multi-petaflop computing facilities for climate prediction, with a ball-park cost of \$1 billion over 5 years. The idea will be presented at the international climate negotiations opening in Bali, Indonesia, this week, with a formal proposal to be published next year in *Bulletin of the American Meteorological Society*.

Integrate and accumulate

The data expected will not just be more precise — they will also be more wide-ranging, providing new impetus to models that seek to treat the Earth system as an integrated whole. Until recently, Earth observation has been less than the sum of its oceanic, terrestrial and atmospheric parts, according to Stephen Briggs, head of science, applications and future technologies at ESA. “Integrating the components is something we are really bad at,” says

Only Germany, however, has so far made a firm long-term funding commitment. In the United States, which contributes 50% of the overall programme costs, funding currently comes through the National Oceanographic and Atmospheric Administration. But it is not yet clear which agency will be responsible for maintaining the array once Argo becomes a routinely operational system. Proof of definite advantages — for fisheries, merchant shipping, oil-spill management or naval purposes — could help secure long-term funding from more governments, says John Gould, a consultant and former director of the project. “But it’s still rather early days to evaluate the benefits.”

Raft of achievements

The first scientific achievements using data from Argo include the detection of accelerated circulation in the subtropical South Pacific¹, the tracking of the deep water that forms in the Labrador Sea² and the observation of local sea-temperature changes during hurricanes. But there have also been setbacks. A surprising suggestion that the oceans’ heat content was diminishing, rather than increasing as expected³, turned out to be an artefact caused by a software glitch⁴.

Gould says that this underlines the need for experienced scientists — who know both the oceans and the technology — to be involved in data management. “We’re still learning,” he says.

Argo data are already incorporated into models for seasonal weather prediction. To initialize such forecasts, scientists ‘tell’ their

Briggs. “This is where we are going to see the major advances.”

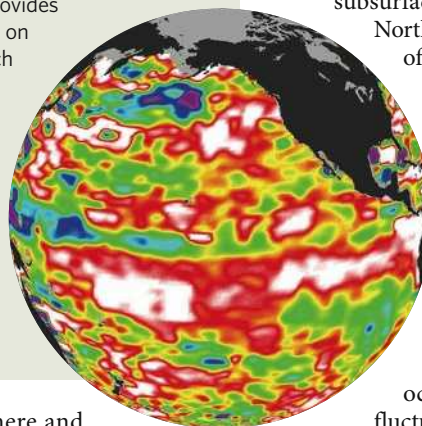
Incorporating more geophysical observations made from multiple instruments obviously makes models more complex. To integrate such disparate data sets, which differ not just in their spatial and temporal resolution but also in their error profile, modellers are borrowing the ‘data assimilation’ techniques used by weather forecasters. A model producing a weather forecast will start off with reasonable best estimates of initial global conditions informed by the data to hand. As more come in — as low-Earth-orbit

The Jason project

Jason is a joint French/US satellite designed to monitor the oceans’ topography. The project follows on from TOPEX/Poseidon, an earlier collaboration between the French space agency CNES and NASA, in which a French and an American radar altimeter shared the same antenna. Jason is an operational continuation of that earlier research effort. Its successor Jason-2 will move yet farther down the road to operational use, bringing in EUMETSAT, the US weather-satellite operation.

Ocean altimetry provides data on sea level and on ocean currents, which deform the surface. There is a significant synergy between Jason’s data on the surface of the ocean and those from the Argo float array on the ocean’s inner structure.

Oliver Morton



models about the here and now and then cast them off into the future. Precise knowledge of the initial state of the ocean — which has a longer ‘memory’ than the atmosphere — could greatly improve the accuracy of longer-term weather and climate prediction. At the UK Met Office’s Hadley Centre in Exeter, Argo data have been used to initialize the centre’s new decadal temperature-forecasting system, which combines observations and models on longer timescales⁵.

“The data are absolutely invaluable,” says Matt Huddleston, a climate scientist at the Hadley Centre. “We now use Argo data continuously for forecasting everything from European winters to tropical storms and El Niño events.” The success of the Met Office in correctly predicting some very cold and wet conditions during the 2005–2006 European

satellites pass over new places, for example — the model’s evolution is iteratively compared with reality. So sparse and infrequent sources of data can still play a role.

Counter-intuitively, more data sources can also often simplify modelling, as they can help to better define other variables, adds Trenberth. For example, raw measurements from a buoy might be misleading if it were in an eddy of the warm Gulf Stream rather than somewhere more representative of the Atlantic as a whole. A system that could use other data to know that the buoy was in the Gulf Stream would not be misled so easily, and the model would be made more realistic. “As one can resolve features better,” he predicts, “one can utilize data better.” The resolution and the data can be provided, if the institutions allow; it will then be up to Trenberth and his colleagues to make good on that forecast.

Declan Butler is a senior reporter for Nature, based in France.

See Editorial, page 761.

winter seems to have been made possible in part by Argo data, which revealed an abnormal subsurface temperature pattern in the North Atlantic. But statistical proof of such benefits will be available only when the Argo network has been maintained for much longer.

Perhaps Argo’s most valuable contribution will be in facilitating studies of year-to-year and decade-to-decade variability in the oceans. At present, it is hard to distinguish climate-change-driven shifts in ocean circulation from natural fluctuations. Argo will provide the continuous time-series needed to solve such puzzles. It should also help settle the big question of the extent to which the oceans participate in climate variability.

“We keep learning that it is dangerous to infer changes in the oceans from only a few years of measurements,” says Brian King, a physical oceanographer at the Southampton Oceanography Centre, UK. “But if anthropogenic forcing does leave a mark on the ocean, Argo should definitely help us find out.”

Quirin Schiermeier

1. Roemmich, D. et al. *J. Phys. Oceanogr.* **37**, 162–173 (2007).
2. Fischer, J. & Schott, F. J. *J. Phys. Oceanogr.* **32**, 573–584 (2002).
3. Lyman, J. M., Willis, J. K. & Johnson, G. C. *Geophys. Res. Lett.* **33**, L18604 (2006).
4. *Nature* **447**, 8–9 (2007).
5. Smith, D. et al. *Science* **317**, 796–799 (2007).

Not enough eyes on the prize

The capacity of the United States to monitor Earth's vital signs is being stymied by tight budgets and poor coordination. **Alexandra Witze** reports.

It seems like such a little thing, the ability to lie back and look up at the full Moon. A moment of wonder or romance on a summer evening, perhaps, but not something vital to the way you do your job. Unless, that is, your job is measuring the amount of photosynthesis going on in Earth's oceans.

SeaWiFS is a NASA satellite that spends most of its days staring down at the ocean, measuring the subtle colour changes that come about as phytoplankton levels wax and wane. But once a month, SeaWiFS takes its electronic eyes off the water, rolls itself backwards and takes a picture of the full Moon. Without the check-up on its colour perception that this regular 'lunar calibration' provides, SeaWiFS could find its judgement drifting slowly off-kilter.

Unfortunately, SeaWiFS can't keep this up for ever. It was designed for five years and has lasted ten. And at the moment, there is no replacement quite as good. Two of NASA's other Earth-observing satellites carry sensors that can measure ocean colour, as do Europe's Envisat mission and some other satellites. But none of them is as good at monitoring ocean colour as SeaWiFS, says the project's chief scientist Gene Carl Feldman of NASA's Goddard Space Flight Center in Greenbelt, Maryland. And the US instrument designed as a direct follow-on to SeaWiFS is just not as good, many say. It has fewer wavelength bands, it might have problems correcting for atmospheric turbulence, its observations could be corrupted by stray light leaking in around the sensors — and it will never take time out to check its colour vision by staring at the Moon.

In other words, the next generation of ocean-colour sensors, built by the most advanced research nation in the world, will in some respects be a step back. And this is not an isolated problem. Climate scientists have a list of a couple of dozen 'essential climate variables' (see 'The dimensions of the problem') that they would wish to see monitored in perpetuity. In fact, they'd like a lot more than that — the list started out with more than 150 variables and was winnowed down in large part on the basis of what data were readily available. "There was a fair amount of pragmatism," says Kevin Trenberth, a climatologist at the National Center for Atmospheric Research in Boulder, Colorado. Nevertheless, in some cases the relevant measurements are not yet being made (see 'The crucial measurement', page 785). And although some data sets are being interrupted or degraded, others are duplicated — provided by multiple satellites and multiple nations.

Atlantic productivity revealed by SeaWiFS.

SEAWIFS PROJECT, NASA/GODDARD SPACE FLIGHT CENTER, ORBIMAGE

THE DIMENSIONS OF THE PROBLEM

Key data that are needed to understand the Earth system: some monitors are listed here.

Atmosphere

Surface wind speed and direction	ERS-2 (Europe), QuikSCAT (US)
Upper-air temperature	Aqua (US), Metop (Europe), GOES series (US)
Water vapour	GOES series (US), Metop (Europe)
Cloud properties	CloudSat (US), CALIPSO (US), Metop (Europe)
Precipitation	TRMM (US/Japan), Aqua (US)
Earth radiation budget	Aqua (US), Meteosat (Europe), SORCE (US)
Ozone	Aura (US), Terra (US), ERS-2 (Europe), Envisat (Europe)
Aerosols	Parasol (France), Envisat (Europe), Terra (US), Aqua (US)
Carbon dioxide, methane and other greenhouse gases	Terra (US), Aura (US), Metop (Europe), Envisat (Europe)
Upper-air wind	Meteosat (Europe), GOES series (US)

Oceans

Sea ice	ERS-2 (Europe), Aqua (US)
Sea level	Jason (US/France), Envisat (Europe)
Sea surface temperature	Aqua (US), Envisat (Europe)
Ocean colour	SeaWiFS (US), Envisat (Europe), Aqua (US)
Sea state	Factors such as roughness, usually monitored from the surface
Ocean salinity	Argo float system, plus other float and ship-towed sources

Terrestrial

Lakes	Landsat (US), Envisat (Europe)
Glaciers, ice caps and ice sheets	IceSat (US), Terra (US), Envisat (Europe), ERS-2 (Europe), RADARSAT (Canada)
Snow cover	NOAA series (US), Terra (US), Envisat (Europe), RADARSAT (Canada)
Albedo	Landsat series (US), SPOT series (France)
Land cover	Landsat (US), Envisat (Europe)
Fraction of incoming solar radiation absorbed by plants	Usually done through small field-based studies, but also Envisat (Europe), Terra (US)
Leaf-area index	Usually done through small field-based studies
Biomass	Terra (US), Landsat (US), Envisat (Europe)
Fire disturbance	Terra (US), GOES series (US), ERS-2 (Europe)
Soil moisture	ALOS (Japan), Landsat (US)

The problems are global, as each nation struggles to fund and maintain data streams from satellites that serve its own interests (which may be shaped by the particular research interests of its scientists). But the issues are most apparent in the US government's civilian Earth-observing satellites, of which there are 30. Turf battles among multiple US agencies, as well as tight budgets, threaten the future of the country's Earth monitoring. The National Academies, not known for alarmist views, was prompted earlier this year to note that "the United States' extraordinary foundation of global observations is at great risk".

The administration sees talk of a crisis as unwarranted. "I think the panic is greatly exaggerated," says Vice-Admiral Conrad Lautenbacher, head of the US National Oceanic and Atmospheric Administration (NOAA), which oversees the country's weather satellites

as well as its fisheries and other ocean resources. Lautenbacher is, among other things, a great believer in putting national capabilities in a broader global context. He has been a driving force in America's contribution to the creation of a worldwide coordinating network for Earth observations, the Global Earth Observation System of Systems (GEOSS), and hopes that such networking can provide the seamless integration of Earth-monitoring systems needed to protect society against natural hazards. GEOSS was launched two years ago with a ten-year mandate to get the world's view of its common house in order. Lautenbacher and other supporters say that although it is moving forward slowly (see 'All in this together'), it has garnered enough political support to address some of the major issues with data gaps and other observational problems.

GEOSS does little, however, to solve one of

the sector's fundamental problems: bridging the long-standing gulf between the scientific community, which generally wants to fly cutting-edge instruments with which to discover things, and the operational community that has the job of providing long-term but unglamorous data sets. In the United States, NASA builds and launches research satellites and NOAA handles the operational systems such as weather satellites. But NOAA does only oceans and atmosphere — keeping track of things on land is the responsibility of the US Geological Survey (USGS), which runs the Landsat Earth-observing satellites with NASA. Depending on what piece of information about Earth is needed, it could have been gathered for any number of purposes by any one of the three agencies. And that's not even counting private remote-sensing spacecraft, nor the military satellites gathering both classified and unclassified data.

United front

One way to simplify things would be to have a unified system for operational measurement of the variables of interest. This is the purpose of Europe's Global Monitoring for Environment and Security (GMES) programme (see page 778). Yet attempts to unify disparate systems can bring problems of their own. The National Polar-Orbiting Operational Environmental Satellite System (NPOESS) combines activities previously carried out by the NASA and NOAA low-Earth-orbiting satellites with the defence department's weather-satellite programme. Begun in 1994 and run by Northrop Grumman, by 2005 the NPOESS had accumulated so many cost overruns that it triggered a mandatory federal review. Its estimated cost of nearly US\$7 billion had soared to at least \$11 billion, and the initial launch date has slipped from 2009 to 2013.

In response, project officials yanked five climate sensors off the NPOESS satellites to save money. This limited their capacity to monitor several of the essential climate variables, including Earth's radiation budget and atmospheric ozone concentrations. Climate scientists protested against the NPOESS cancellations, and the National Research Council is now conducting a review to see how the data from the sensors might be obtained through other means; its report is expected in January 2008. Options include mounting some of the sensors on other spacecraft or reinstating them later in the NPOESS series and just making the best of the gaps thus created. "The plan isn't going to be as complete as we would like it to be, but it will be as complete as we can be," says John Marburger, the science adviser to President George W. Bush.

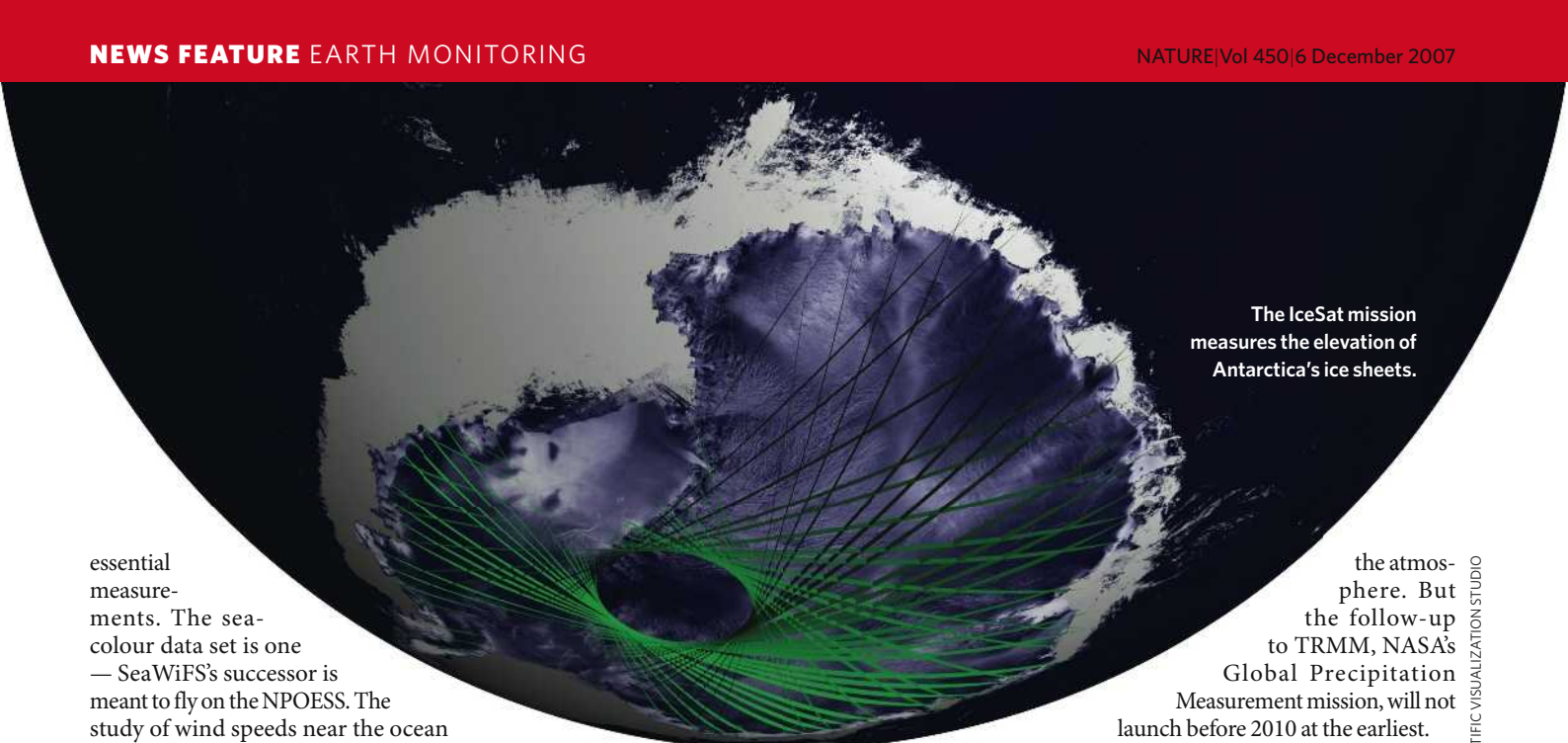
Delays to the NPOESS threaten various

essential measurements. The sea-colour data set is one — SeaWiFS's successor is meant to fly on the NPOESS. The study of wind speeds near the ocean surface is another. A microwave radiometer that was supposed to make such measurements has been removed from the first NPOESS satellite scheduled to launch to the second. The US instrument currently in orbit doing that task — NASA's QuikSCAT — is long past the end of its design lifetime (and was itself a last-minute replacement thrown up into orbit after an earlier scatterometer failed after less than a year). QuikSCAT measurements are regularly used to improve hurricane forecasts, and Bill Proenza lost his job as director of the National Hurricane Center in Miami in July partly because he criticized the lack of plans to replace the satellite (see *Nature* **447**, 514–515; 2007).

There is another option for measuring wind vectors, which is the WindSat instrument currently flying aboard the joint military–civilian Coriolis satellite. But that — like the planned replacement to come aboard the NPOESS — is

a passive radiometer, measuring microwaves that are emitted from the sea surface rather than actively bouncing microwaves off the ocean and observing them. Researchers continue to debate whether a passive radiometer can measure winds as accurately as an active scatterometer.

There is also bad news for satellites that have nothing to do with the NPOESS programme. The US/Japan Tropical Rainfall Measuring Mission (TRMM) provides detailed images and data such as the amount of rain produced from hurricanes and other tropical storms. Its nominal three-year mission came to an end in 2000, but it has had its lifetime extended again and again, most recently until September 2009. The latest extension meant that NASA had to waive a safety requirement that it maintain enough fuel to ensure that it burned up on re-entry into



The IceSat mission measures the elevation of Antarctica's ice sheets.

the atmosphere. But the follow-up to TRMM, NASA's Global Precipitation Measurement mission, will not launch before 2010 at the earliest.

Gaps between key missions are a long-standing problem in Earth observation, particularly for those looking for long, complete data sets. If there had not been a gap of ten years between SeaWiFS and the instrument it replaced, scientists would have a far better understanding of how ocean productivity changes with weather and climate; there would be more data on how different El Niño events and other fluctuations in ocean temperature control phytoplankton blooms — work that could provide hints as to how future climate change could affect oceanic productivity (M. J. Behrenfeld *et al.* *Nature* **444**, 752–755; 2006). Such gaps are particular problems for data from research instruments that have not yet been put on an operational basis — but it can afflict operational systems too.

The Landsat series of satellites, for instance, has been monitoring Earth constantly since 1972, and one of its main purposes is to provide a continuous stream of data so that changes — such as urban growth, land subsidence and desertification — can be measured over time. But its issues with data continuity sprang to the fore in early October, when Landsat 5 — a satellite that had been up since 1984 — stopped working.

Landsat 5 was to have been replaced by Landsat 6, but that satellite failed on launch in 1993. Landsat 7, launched in 1999, has a problem with its scanning mechanism that causes it to collect data in zigzag streams rather than continuous bands, which leads to about 22% of each scene being lost. The USGS continues to process and release usable data from Landsat 7, but no one can now be sure whether it will last until the launch of the optimistically named Landsat Data Continuity Mission, an event currently scheduled for 2011. Programmatic discontinuities haven't helped; over the years, the Landsat programme has been passed between NASA, NOAA, the USGS and a private operator like a cold around a kindergarten.

Although data gaps remain a major issue,

All in this together

Last week in Cape Town, ministers of more than 70 countries gathered to talk about the biggest acronym in Earth sciences: GEOSS, the Global Earth-Observing System of Systems.

Touted as a worldwide network of ocean, atmospheric and terrestrial sensors, GEOSS is supposed to hook the planet together in one big harmonious Earth-monitoring whole. Two years into its ten-year implementation plan, the project is already bragging about its 'first 100 steps' towards that goal.

These include GeoNetCast, a web portal to broadcast information from Earth-monitoring systems

around the world; SERVIR, a programme to help Central America improve its monitoring of forest fires and tropical storms; and the agreement of a Brazilian–Chinese venture to share satellite data with Africa for free. All showcase the main point of GEOSS, which is to serve nine 'societal benefit areas', such as protecting water and energy resources and reducing deaths caused by natural disasters.

The societal focus of GEOSS has turned off some scientists who had hoped it might provide new funding streams for their remote-sensing work, says José Achache, the secretariat of the Group on

Earth Observations in Geneva, Switzerland, the body that oversees GEOSS. And it's not uncommon to hear grumbling about the programme, even among climate scientists. "It has never reached anything like its promise," says Kevin Trenberth of the National Center for Atmospheric Research in Boulder, Colorado.

Still, many say that GEOSS has managed to at least focus interest at the ministerial level on Earth monitoring, and as such has raised the political profile of the Earth sciences. And at the very least, its existence has prompted member countries to get their own Earth-monitoring houses in order.

A.W.

overlap between monitoring systems is another. Frequently the problem is international; one country launches a spacecraft that partially duplicates what another mission is already doing. International steering committees are supposed to cut down on the overlap, but it doesn't always work that way. "Our hope is not just to fill gaps, but to avoid duplication of effort," says Helen Wood, a senior adviser to NOAA's satellite and information services division in Silver Spring, Maryland.

Poles apart

In 2003, NASA launched its ICESat mission mainly to study the ice sheets of Greenland and Antarctica; in 2005, the European Space Agency launched its CryoSat, which was to have done much the same thing (although it would also have measured sea-ice thickness). CryoSat failed on launch, so now plans are under way to send up a second version in 2009. Meanwhile, NASA — which likes its mission so much, despite a premature laser failure, that it renamed one of its streets at the Goddard center as ICESat Road — is looking at launching an ICESat-II. Waleed Abdalati, who is ICESat's programme scientist, says that so many changes are expected in the Arctic that both the US and European missions will be useful

in providing more information.

José Achache, secretariat of the Group on Earth Observations in Geneva, Switzerland, is not so sure that duplication is a good way forward. "Essentially the agencies were in unofficial competition," he says. An international steering group, the Committee on Earth Observation Satellites, exists to try to cut down on duplication for satellite-based systems, but sometimes national interests win out. The European Space Agency, for instance, is planning a Soil Moisture and Ocean Salinity mission — measuring two of the essential climate variables — at

the same time that NASA had been planning the Hydrosphere State (Hydros) mission for soil moisture — which has since been put on indefinite hold — and the Aquarius mission for ocean salinity.

Sometimes, though, the US–European competition can work in science's favour. With SeaWiFS possibly close to dying, NASA is looking at how it can jump in on the European MERIS instrument, aboard Envisat, to get ocean-colour data, says Paula Bontempi of NASA headquarters in Washington DC. Although the data may not be all that the scientists wish they were, they will be better than nothing once SeaWiFS gives out.

And in the long run Europe plans to have an instrument as good as or better than MERIS as part of the Sentinel 3 series of operational climate-monitoring satellites. Duplication of efforts is undeniably wasteful in the Earth-monitoring world. But relying on any single nation, even the richest and most technically advanced, would risk dooming the planet to an endlessly repeated history of research satellites operating long after their intended lifespans, last-minute scrambles to keep things going, and possibly catastrophic gaps.

Alexandra Witze is *Nature's* chief of correspondents for America. See Editorial, page 761.



The crucial measurement

The "Carbon Club" began meeting on Fridays about a decade ago, setting up shop in whatever spare meeting places it could find at the Jet Propulsion Laboratory in Pasadena, California. Its members, a handful of scientists with extensive experience in remote sensing of Earth's atmosphere, set about brainstorming ways to provide one of the most crucial data sets of the twenty-first century: precise measurements of carbon dioxide levels in the atmosphere on a fine enough scale to definitively track the gases' sources and sinks. "No one was crazy enough to say that they could do it until we came up with a possible solution," says Charles Miller of the Jet Propulsion Laboratory, and one of the original Carbon Club members. That solution is due to reach orbit late next year or early the year after in the form

of a US\$300-million-or-so satellite called the Orbiting Carbon Observatory (OCO).

When the clock starts ticking on the Kyoto Protocol's five-year commitment in January 2008, developed nations that have ratified the treaty will be bound to a strict bookkeeping system for greenhouse-gas emissions. They will receive credits for mopping up their emissions with so-called carbon 'sinks', such as through reforestation efforts and improved agriculture and grazing practices. Yet it is currently impossible to pinpoint where the gases originate — and no one really knows where they end up. Half the CO₂ pumped into the atmosphere by burning fossil fuels ends up in the oceans or absorbed by plants on land — but how much goes each way, and precisely where, is still unclear. "Certain people will tell you emphatically that it's going into the oceans, and they

think they know roughly where it is going in," says Ross Salawitch, an atmospheric chemist at the University of Maryland in College Park, a member of the OCO team and another Carbon Club veteran. "Others will tell you emphatically that land is taking up the carbon. There's nowhere close to a unanimous opinion."

As it orbits Earth, OCO will measure the 'fingerprint' that CO₂ leaves in the air between the satellite and Earth's surface almost half a million times a day. The resulting map of CO₂ concentrations will then be used, with other data and modelling, to work out where CO₂ is being emitted and absorbed. "It's the most difficult atmospheric trace-gas measurement that's ever been made from space," Miller says.

If OCO's two-year mission is a success, it could well serve as a model for an operational mission that might be tied directly to a

A. MARTIN

post-Kyoto regulatory system. But the team is keen not to offer operational data too early. "A prototype always produces challenges," says David Crisp, OCO principal investigator and a senior research scientist at the Jet Propulsion Laboratory.

Today, a network of ground-based stations strung across the globe measures CO₂ and other greenhouse gases at Earth's surface with high precision, but patchy coverage (see page 789). Large expanses of Earth, including Africa, India, Siberia and much of South America, have very few, if any, monitoring stations — even North America, which hosts the highest concentration of measurement stations, has significant gaps in its coverage — the Yukon, for example, and large chunks of Quebec and the US southwest. What's more, the network was specifically designed to avoid picking up the fluxes that OCO is interested in. "The network was actually sited as far away from known sources and sinks of CO₂ as possible so that we could get good, clean, average measurements," says Crisp.

Bounce back

OCO came into being in 2001, when NASA set up a competition for low-cost Earth-science missions. Thirty-three proposals went in: OCO came out. Its instrument works by measuring visible and near-infrared sunlight that is reflected back from Earth's surface — sunlight that has travelled through the atmosphere twice, once going down and once returning up. As sunlight shines down and is reflected back, various molecules absorb some of it at distinctive wavelengths. By comparing the different bands associated with CO₂ and with other gases (which serve as calibrations), the instrument comes up with an estimate of the number of CO₂ molecules in a column of air just 10 kilometres in cross-section. Feed these data, which have much higher resolution than those obtained in previous efforts, into models of atmospheric circulation and you can work out how and from where the gas is spreading.

The OCO team is attempting to measure differences in trace gases with a 1 part per million precision against a background of 380 parts per million of CO₂ equivalent (the approximate concentration of CO₂ in the atmosphere today) while the spacecraft travels at 7 kilometres per second. Although measurements of trace gases on Earth and even on Mars are made down to parts per billion, the interest there is in absolute levels, not in small changes. "If we measure two parts per million more CO₂ over the eastern part of the country versus the western part, that



Do look down: the Orbiting Carbon Observatory could provide precise data about the origin of carbon emissions.

has gigantic implications for the carbon sinks that we would infer," Salawitch says. "Our whole science is driven by small spatial gradients in the gas."

But the measurements face various problems. For instance, they rely on sunlight, so they can't be made at night, or during polar winters. This is a "fundamental shortcoming" says Berrien Moore, a mathematician at the University of New Hampshire in Durham, who studies the carbon cycle. The OCO team has argued that this is not terribly problematic (C. E. Miller *et al. J. Geophys. Res.* **112**, D10314; 2007). A bigger concern for them, reflected in the instrument's



"This is fundamentally a science experiment." — David Crisp

design, is cloud cover. The instrument has a very small viewing window because a thin column of air is less likely to be beset by clouds than is a broader swath. This should mean that the instrument gets enough data to do its job — but places that are frequently cloudy, such as the Amazon rainforest, could still prove troublesome, says Pieter Tans, a senior scientist with the National Oceanic and Atmospheric Administration in Boulder, Colorado, who provides the ground-based measurements that will be used as a validation standard for OCO. Aerosols, too, are a potential problem, says Moore. But he's still excited about the mission: "It is potentially going to be

a huge breakthrough on this source/sink problem, and it will be a terrific pathfinder."

Dynamic duo

A Japanese satellite named Greenhouse gases Observing Satellite, or GOSAT, is scheduled for launch in August 2008, and will complement OCO. GOSAT will measure methane, water and ozone as well as CO₂. Whereas OCO uses a spectrometer based on diffraction gratings, which achieves a high signal-to-noise ratio and thus a precise determination of levels, GOSAT will obtain its measurements with a spectrometer that operates at both short and long infrared wavelengths. Long wavelengths allow it to measure emissions even when there is no sunlight, avoiding the issues associated with night-time or polar winters. And whereas OCO will make spatially contiguous measurements along a narrow field of view (10 kilometres) over a 16-day cycle, GOSAT will measure isolated footprints of the gases over a broad (up to 900-kilometre) swath that repeats every 3 days. "It is a tremendous advantage to the global carbon-cycle community that both approaches are being used during the flagship missions," Salawitch says.

So far, OCO has glided through its first testing stage inside a thermal vacuum chamber with no insurmountable problems. The team is getting ready to test the performance of the instrument early in 2008. But every day between now and launch poses challenges, says Crisp. Perhaps to remind himself as much as anyone else, Crisp says, "This is fundamentally a science experiment. We're asking whether this technique will work as well as the models are telling us it will." With Kyoto taking effect and a reliable bookkeeping system for carbon sources and sinks sorely needed, it's not just the scientists who will be awaiting the results. ■

Amanda Haag is a science writer in Colorado.

See Commentary, page 789.

OCO/JPL/NASA

Thought: a different perspective

SIR — Patricia Churchland's review of Steven Pinker's latest book *The Stuff of Thought* ('Poetry in motion' *Nature* **450**, 29–30; 2007) offers scant information about the book, and what there is incorrect. Churchland instead presents her own views on how molecular biology and neurobiology provide challenges to Pinker, but in so doing she undermines the successes of these disciplines. She concludes that Pinker's book is only about semantics and that his discussion of the mind represents a kind of madman nativist perspective, ignoring the role of the environment and research in the neurosciences.

I have the impression that Churchland restricted her reading to the prologue, heaving the book across the room in dismay while ejaculating "Same old, same old!" Otherwise, she would surely have come across Pinker's detailed analysis of the evolution and development of the core conceptual structures of space, time, number and cause, and how these building-blocks enable the child to acquire not only a lexicon, but also an understanding of the world. This view doesn't eliminate either experience or cultural processes, but rather shows ways in which a core architecture may constrain the acquisition of knowledge and lead to a suite of shared mental capacities.

She would also have come across rich and entertaining chapters on naming our children, swearing (I refrain), and the pragmatics of bargains, bribes and other social conventions. And throughout, Pinker mentions work in the neurosciences. This includes studies of people with brain damage, cellular recordings of animals and humans, and imaging experiments; some of these Pinker conducted himself with colleagues and students.

But these ideas are sometimes controversial, and it saddened me that the Book Review did not discuss why. Instead, it went into challenges apparently posed by genetics and neurobiology for the cognitive sciences, and particularly for the brain-as-merely-hardware kind that Pinker is said to peddle.

Take, for example, Churchland's assertion that "extravagant claims about human uniqueness must deal with the discovery that humans have only about 28,000 genes, and differ from mice in just 300 or so". (Not so: humans have only about 300 genes not found in mice, but the others aren't identical.) Even if accurate, this would not constitute an argument against any of Pinker's book. In fact, it shows why one has to be careful in interpreting the relationship between genomic sequence overlap and phenotypic similarity. The monumental cognitive gap between mice and humans tells us that the number of homologous genes and the

percentage of sequence overlap are simplistic measures of species similarity, rather than the genomic overlap telling us that humans are cognitively equivalent to mice.

We have to look to another story to explain how, given such overlap, we are so different. The point is magnified when we consider the 98% overlap with chimpanzees, and again, the spectacular differences in our cognitive abilities, ranging from the expression of language, music and mathematics, to the creation of soufflés, Global Positioning System navigators and humour. I hope the readers of *Nature* will dig into Pinker's book, even if it is only to learn what he said.

Marc D. Hauser

Department of Psychology, Harvard University,
33 Kirkland Street, Cambridge,
Massachusetts 02138, USA

Thought: book review has my ideas back to front

SIR — Patricia Churchland's review of my book *The Stuff of Thought* ('Poetry in motion' *Nature* **450**, 29–30; 2007) says virtually nothing about the book's contents, and gets two of its main claims backwards. A lengthy section of the book argues against the idea that "thought is like external language in all important respects." And the theory of Jerry Fodor's that Churchland calls "font-change semantics" (whereby a person's knowledge of the meaning of a word, such as *cut*, consists of a single mental symbol, such as 'cut') is one that I argue against, together with Fodor's innateness *ad libitum* claim, also mistakenly attributed to me.

The book apparently stimulated the reviewer to free-associate to her own beliefs that psychological phenomena can be explained at the level of neurons and that human thinking is in the service of motor control. The fact that I (like most cognitive psychologists) have not signed up to these views is the only point of contact between my book and her review.

Steven Pinker

Department of Psychology, Harvard University,
33 Kirkland Street, Cambridge,
Massachusetts 02138, USA

Chance to learn and teach in the developing world

SIR — As married academics on sabbatical leave from the University of California, Davis, we would like to add some comments to your Naturejobs feature on the subject ('The seven-year itch' *Nature* **448**, 834–835; 2007). We felt that you overlooked the importance of academics being role models for people in the developing world, where

education is often less than perfect, and meeting the challenge of solving problems in the developing world, where medical, poverty and conservation issues often reach their zenith.

We have moved from California to Dar es Salaam, Tanzania, for our sabbaticals. Here we can interact with Tanzanian academics, work with local non-governmental organizations, use the Internet to write a book, and continue our field work. We hope to expand our intellectual horizons and those of Tanzanian colleagues, and perhaps bring fresh perspectives to developing-world problems. Broadening academic influence outside first-world universities is an important social responsibility. We urge other sabbatical wannabes to take up similar challenges and opportunities.

Tim Caro*, Monique Borgerhoff Mulder†

*Department of Wildlife, Fish and Conservation Biology, University of California, Davis,

†Department of Anthropology, University of California, Davis, California 95616, USA

Few women join ranks of Germany's academic elite

SIR — The excellence initiative is indeed a welcome addition to the German academic world, as your Naturejobs Feature 'Allowing an elite' points out (*Nature* **450**, 452–453; 2007). But you omit to mention the plight of women academics. Only 10% of full professors at German universities are female. Just 14.6% of all grants from the DFG, Germany's main funding agency for university research, were awarded to women in 2006, a mere 2% rise since 2003.

I would be interested to know how many women received financial 'excellence cluster' awards, and whether the DFG is making attempts to reward female scientists in Germany's male-dominated academia. The Feature, whose interviewees and examples were all male, conveys the impression that this will continue to be a man's world.

Laura Niven

Department of Human Evolution,
Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany

The German government is keen to increase the number of female scientists in top positions. On 19 November, the federal science minister and the Länder (state) governments agreed on a €75-million (US\$111-million) programme for creating 200 new tenured positions for female university professors in the next five years — Editor, *Nature*.

Contributions to Correspondence may be submitted to correspondence@nature.com.

COMMENTARY

Cinderella science

On-the-ground monitoring is unglamorous work, seldom rewarded by funding agencies or the science community. But we neglect it at our peril, warns **Euan Nisbet**.

Sometimes discovery comes slowly, not with a flash of revelation but creepingly, as larger patterns emerge painfully from years of data. Researchers who work in Mauna Loa, Hawaii, are celebrating the fiftieth anniversary of a measurement programme responsible for the longest continuous recording of atmospheric carbon dioxide. Seldom can such insight have grown (and continued to grow) over so long a time. Now that we look back, the resulting 'Keeling curve' of CO₂ concentrations ranks very high indeed among the achievements of twentieth-century science.

Charles David Keeling's account of his tribulations¹, "Rewards and penalties of monitoring the Earth", should be compulsory reading for politicians and science administrators. Idealistic young scientists, as yet unscarred, should read it and take note: courage and perseverance are required. Before Keeling, little was known about CO₂ in the atmosphere and available measurements had little value. Success came from Keeling's painstaking years of effort and innovation. Despite the import of the results, the work was often threatened, as is attested by a gap in 1964 when underfunding briefly halted measurement.

Monitoring is science's Cinderella, unloved and poorly paid. Sustaining a long-term, ground-based programme that demands high analytical standards remains challenging. Funding agencies are seduced either by 'pure' notions of basic science as hypothesis-testing, or by the satanic mills of commercial reward. Neither motive fosters 'dull' monitoring because meeting severe analytical demands is not seen as a worthwhile investment. At one stage, Keeling was ordered to guarantee two discoveries per year and today, modern research has become a planned journey through set 'milestones' to deliverable destinations.

No longer do we blindly cast our bread on the ocean of truth. Keeling's long-term CO₂ measurements began in 1957 with the first flask collection at the South Pole. Hawaiian measurements started in March 1958. That air had 316 parts per million of CO₂. By March 2007 the comparable value was 384 parts per million. As data curves lengthened, patterns emerged. Seasonal changes and hemispheric differences traced the breathing of the biosphere, dominated by springtime CO₂ uptake and autumn

"The Mauna Loa curve thrust itself before humanity's eyes, changing our view of the world."



Charles David Keeling's work made us aware of rising amounts of carbon dioxide in the atmosphere.

release in the Northern Hemisphere. Keeling measured the isotopic abundance of carbon-13 in CO₂ to show that the seasonal changes were caused by land plants.

Keeling's first report is a landmark², documenting the seasonal cycle and, more gloomily, the annual rise in CO₂. By the 1970s, the biological influence of the El Niño/Southern Oscillation became clear as climate shifts altered the seasonal CO₂ cycle. And the connection between rising CO₂ and fossil-fuel burning was firmly established, showing that a substantial fraction of CO₂ added by humans remains in the atmosphere and is not removed by the biosphere.

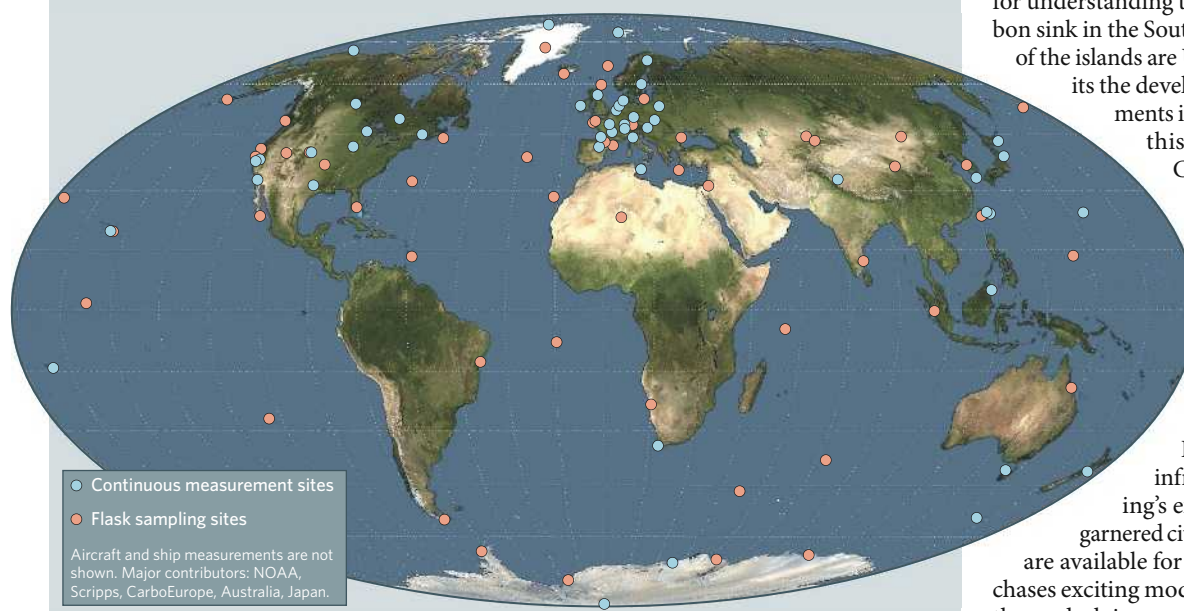
The Mauna Loa curve, simple and unambiguous, thrust itself before humanity's eyes, changing our view of the world. Keeling's work was far ahead of its time. It was the 1970s before other quality-controlled data sets got going. Had we not had his long back-record, awareness of global change would have come more slowly. Sudden events, such as the marked fluctuations in global CO₂ uptake after the 1991 volcanic eruption of Mount Pinatubo, may have looked very different in the context of a 15-year rather than a 30-year record.

Global understanding is underpinned by rigorous *in situ* (measured on the surface) long-term data series. Since 1957, satellites have patrolled the skies, and they are crucial for the wider picture. But they need to work in partnership with the much more accurate *in situ* monitoring. Satellite monitoring originally missed the ozone hole, found by careful ground measurements. Following Keeling's example, *in situ* measurement spread across the globe. The World Meteorological Organization (WMO) set up the Global Atmosphere Watch, widening the coverage of the US National Oceanic and Atmospheric Administration (NOAA) network, to monitor long-lived CO₂, methane, nitrous oxide and other greenhouse gases on a global scale.

In it together

Nearly all the G8 powers, plus China, Brazil and Indonesia, share in the WMO network. Dozens of countries help, from Kenya to Kyrgyzia. Far-flung huts hum with recording instruments — from the Cape of Good Hope, to Alert in Arctic Canada, Zeppelin in Spitsbergen, Norway, and Cape Grim in Tasmania. Supporting the continuous measurements, groups regularly collect air flasks at many locations worldwide. Guiding the work is the

SIO ARCHIVES, UNIV. CALIFORNIA, SAN DIEGO

THE WORLD'S CO₂ MEASURING STATIONS

expert panel on CO₂ and trace gases, the inheritor of Keeling's enthusiasm. This is the United Nations at its very best, acting for humanity.

"Trust, but verify," said Ronald Reagan of the 1987 US–Soviet nuclear-missile treaty. International agreements such as the Kyoto Protocol can succeed only when verified. Unless independently audited, self-declared emissions reported by individual nations cannot be trusted. Without trust, any Kyoto follow-up treaty surely cannot succeed. Here, measurement promises much. Projects such as NOAA's CarbonTracker, and CarboEurope run by the European Union (EU), offer hope of real validation. Long records of concentration and isotopic data, when modelled with meteorological information, should eventually enable full understanding of global and regional carbon-gas budgets, by location, season and by source or sink type.

The devil is in the detail

As we face an uncertain planetary destiny, modellers seek to interpret the past and predict the future. But comprehension requires good *in situ* data — quality is crucial if we are to understand the long-term increase in greenhouse gases in the air by computer modelling. To achieve deep insights from both data and models, curiosity-driven scientists need to work closely with analysts and technicians who can build innovative instrumentation. Without trustworthy measurement, we cannot understand the system of Earth.

And the gaps remain huge, especially on land in the tropics, the heart of the biosphere (see map). Southern Asia, immensely important to both land biosphere and world economy, is barely measured. Africa and South America have few tropical stations. Isotopic records of carbon gases, potent in identifying sources, are nowhere continuous. Measuring oxygen concentrations provides insight into

biosphere uptake of CO₂, but the monitoring network is small. Nitrous oxide, an easy target for greenhouse reduction but needing precise measurement, is monitored in too few places. Global cooperation is hindered by EU rules on the safety testing of gas cylinders.

The problems faced by Keeling have not gone away. Above all, data sets need to be long-term, sustained over the decades. It is easier to get big funding for one-off projects than small funding for long-term infrastructure. Although *in situ* monitoring is cheap, stations frequently face closure and vital time series are broken. Sustaining continuity and quality is hard. A few years ago, Australia's outstanding programme — now recovering — was savaged by budget cuts. Although German help maintained former Australian monitoring in Scotland, the important carbon-13 time series at Cape Rama on the Indian coast, tracking the monsoon winds, ended. Australia's withdrawal leaves a large gap in southern Asia, at a time when regional efforts need to increase.

In Europe, although Brussels supports projects such as CarboEurope and Geomon, there is no long-term funding commitment. For example, the EU's methane-monitoring network was terminated, with responsibility left to individual nations, although the winds heed no national borders. Satellite infrared monitoring helps, but, needing sunlight, it is blind to methane emissions at night. A better EU policy would support close partnership, not competition, between *in situ* and space observations.

Uniquely among the major nations, Britain does not support atmospheric CO₂ monitoring. To fill the gap, the EU funds measurements in Britain and marine CO₂ work, the United States collects from remote UK islands and France monitors the North Atlantic background levels.

However, the UK absence has consequences for understanding the changing oceanic carbon sink in the South Atlantic — where most of the islands are UK territories — and limits the development of new measurements in the tropics, even though this is a perfect role for the Commonwealth.

Many of Keeling's problems came from the views of the scientific community itself. *In situ* work promises neither shiny rockets nor lucrative contracts. Monitoring does not win glittering prizes. Publication is difficult, infrequent and unread. Keeling's extraordinary 1960 paper garnered citations slowly. When funds are available for climate research, money chases exciting modelling, whereas funds for the underlying monitoring data have to be boot-legged. Modellers are very supportive, recognizing their dependence on gathering true data, but funding committees are not. Despite Keeling's successes, long-term measurement is simply not valued as 'discovery' science.

There is hope. Driven by Keeling's example, the world's unlikely hero — considering its global-warming stance — is the United States. The Scripps Institution of Oceanography in La Jolla, California, where Keeling's son Ralph now works, Princeton University in New Jersey and NOAA deserve praise. Many other nations sustain similar high-quality measurement and Brazil and other concerned tropical countries are starting to make contributions. But moves in Brussels to create an integrated European carbon observing system as a matching partner for NOAA need support. Southern Asia, now becoming wealthy, should do much more, as should the Middle Eastern nations.

Despite the importance of his discovery, Keeling, who died in 2005, was not awarded a Nobel prize. A better monument to his work would be an improved UN Global Atmosphere Watch for greenhouse gases. It is time to build on that, to gain stronger, more permanent commitment by many nations, and to sustain and expand the global *in situ* measurement network, especially in the tropics. The reward will be global security through climate knowledge. With precise, quality-controlled data, we shall better comprehend the Earth system and the breathing of the planet, watch the greenhouse, verify Kyoto, and gain warning of dangers to come.

Euan Nisbet is in the Atmospheric Group, Department of Geology, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK.

"Monitoring does not win glittering prizes. Publication is difficult, infrequent and unread."

1. Keeling, C. D. *Annu. Rev. Energy Environ.* **23**, 25–82 (1998).
2. Keeling, C. D. *Tellus* **12**, 200–203 (1960).

SOURCE: A. C. MANNING, UNIV. EAST ANGLIA

COMMENTARY

Vigilance is not enough

Global surveillance is key to tracking potential pandemic viruses such as H5N1. But we need to share samples more rapidly, increase testing in endemic areas and track more than one virus, argues **Walter Boyce**.

Another influenza pandemic seems inevitable, and without a generic vaccine, our best chance of being prepared is to identify, track and stop the spread of viruses such as highly pathogenic H5N1. Two years ago, some believed that H5N1 viruses were poised to spread around the globe on the wings of migrating wild birds. A massive effort was mounted to track their movement but, as of September 2007, very few positive birds have been found in tests of over 300,000 healthy wild birds from more than 40 countries¹. Several hundred infected birds (almost all of them dead) were found in endemic and outlying areas, but dead birds do not tell us about the birds that don't get sick when infected — those that could spread H5N1 over longer distances.

We shouldn't be surprised that random sampling failed to detect what is apparently a rare event: active H5N1 infection in migrating, healthy wild birds. And failing to find H5N1 can be good news if the goal is early detection in non-endemic areas, as it was for many of the wild birds sampled in Europe and the United States.

But it is troubling that we still don't understand the importance of wild birds in the replication and spread of H5N1 within and between countries and continents. Each new virus outbreak, such as that in free-range poultry in the United Kingdom last month, is followed by uncertainty and speculation about the source and transmission route: is it wild birds or poultry? Knowing the source won't change the immediate response — new outbreaks, like fires, must be stamped out — but it will guide efforts aimed at preventing further outbreaks. For example, moving free-range poultry indoors to separate them from wild birds makes sense if wild birds are a risk. But keeping poultry indoors won't prevent an outbreak if the transfer of infected poultry is responsible.

There is growing recognition that it is time to change our surveillance approach to H5N1 and the other influenza subtypes that could cause the next pandemic. Here are three steps that can move us forward.

First, we must improve our ability and willingness to rapidly share data and samples. Despite calls for a more open approach by researchers such as Ilaria Capua at the National



No access: only by sharing data from local bird-flu outbreaks can global agencies prepare for a pandemic.

Reference Laboratory for Avian Influenza and Newcastle Disease in Padova, Italy, there are still substantial obstacles to the rapid sharing of data and samples, whether they are from wild birds or poultry. Regulatory problems hinder shipping of these samples between countries — it can take weeks to months to arrange for the proper import/export permits, and some countries do not allow any samples to be analysed outside their borders. It is also difficult to balance rapid public release of data with appropriate protection of intellectual-property rights.

But attitudes are changing, as shown by the creation of the Global Initiative on Sharing Avian Influenza Data (GISAID), an organization designed to unite researchers and promote data sharing. In 2005, the World Organisation for Animal Health (OIE) and the Food and Agriculture Organization (FAO) created the OFFLU avian influenza network to facilitate the collection and exchange of viruses and the deposition of sequences in genome banks. And in the United States, the National Institute of Allergy and Infectious Diseases (NIAID) has an established policy of releasing influenza genome data to GenBank and other public sites within 45 days of being generated.

On the wild-bird front, the US Agency for

International Development and the Centers for Disease Control and Prevention have entered into a unique public-private partnership with the Wildlife Conservation Society, a non-governmental organization now charged with managing the Global Avian Influenza Network for Surveillance (GAINS). This initiative aims to establish a global collaborative network for sharing data, including all influenza subtype samples, through an open-access database available to anyone².

Swapping data

Although these and other approaches and policies are important, it is ultimately up to the researchers on the front lines to ensure that data and samples make it into the pipeline as quickly as possible. Because isolating viruses is such hard work and because the resulting samples can yield a wealth of information and publications over time, there may be a tendency to view viruses in the freezer as money in the bank — a rich resource to be guarded and tapped later. This approach does not enhance pandemic preparedness, and future publications won't seem so important in the middle of a roaring pandemic. We simply must do a better job, and making data release within 45 days a community standard would be a step in the right direction.

Second, we must unravel the role of wild-bird

"Future publications won't seem so important in the middle of a roaring pandemic."

K. WIGGLESWORTH/AP



Live birds pose a greater threat of transmitting the flu virus than dead birds (inset).

species in spreading H5N1 by improving our wild-bird surveillance and research efforts. It is clear that wild aquatic birds are a natural reservoir of influenza viruses, although it is possible that H5N1 persists in a wild-bird species we wouldn't normally suspect. But we don't know whether H5N1 viruses are endemic in wild-bird populations or if infections in wild birds represent spillover from poultry. And although movement of infected poultry plays a pivotal part in spreading H5N1 to new areas, we continue to debate, year after year, the risk to poultry posed by migrating wild birds.

To move beyond debate and come up with definitive answers, we must investigate H5N1 in the areas where virus transmission actually takes place. Until recently, most wild birds tested for H5N1 had been sampled in regions of Europe and North America where infections are rare or absent. Dead birds, such as swans, have been conspicuous sentinels of H5N1 infection in Europe, triggering rapid responses to prevent the spread to poultry. But we don't know how these birds acquired their infections, and the presence of infected dead birds doesn't necessarily mean that virus transmission is occurring, or will occur, in the area where they are found.

A live problem

It is live birds, not dead birds, that shed the virus and pose a threat. Because our surveillance efforts have struggled to find infected live birds in non-endemic regions, it is essential that we shift sufficient resources and efforts to evaluate H5N1 transmission in wild birds in known endemic areas such as southeastern Asia, China, Indonesia and Africa.

This is not a trivial task, but encouraging signs are emerging. For example, the FAO in collaboration with national veterinary serv-

ices and wildlife institutions launched a wild-bird surveillance programme in 2006 in Eastern Europe, the Middle East and Africa³. And the GAINS programme, involving dozens of governmental, non-governmental, as well as academic partners, is now working in almost 30 countries, including Cambodia, Mongolia, Vietnam and Indonesia.

Early detection programmes in non-endemic regions are valuable and must continue alongside efforts in endemic areas. But when the data are mostly negative, it is no simple matter to develop a systematic, science-based programme. Risk assessments, such as those performed in the United Kingdom by the Department for Environment, Food and Rural Affairs⁴, help improve surveillance efforts because they specifically consider the ecology and movements of high-risk bird species relative to known outbreaks of H5N1. These assessments, and the value of wild-bird surveillance in general, could be markedly improved by sampling birds at the beginning and end of their migratory journeys from endemic areas.

Third, we need to improve our ability to diagnose and characterize all virus subtypes from wild birds. H5N1 viruses are not the only pandemic threat, and this provides a compelling reason to identify and track the movement of other influenza viruses as well. We need vaccines that protect us against all of the influenza subtypes that nature might throw in our direction. The global human population is immunologically naive to at least 12 viral subtypes in addition to H5, and we don't know which virus will cause the next pandemic, or the one after that. Besides H5, today's 'short list' should include H2, H6, H7 and H9. We must look beyond H5N1 and conduct surveillance that

captures the full list of viral genetic diversity — and feed those data into the development of effective pandemic vaccines.

We don't have to look far. Depending on species, location and season, up to 25% of the wild birds sampled in the past two years were infected with non-H5N1 influenza viruses. The early detection efforts in North America and Europe may have failed to find H5N1, but they did an excellent job of sampling other influenza viruses. Unfortunately, too many of these viruses were discarded and their information lost forever once samples were classified as non-H5N1. This is

not good enough. We must evaluate the pandemic potential of all the influenza subtypes detected during surveillance.

It won't be easy or inexpensive. The OIE has only seven dedicated laboratories worldwide that diagnose H5N1 and other avian influenza viruses, with additional testing done in some national and private laboratories. Unfortunately,

our ability to detect and identify the full range of viruses is limited because the diagnostic reagents and procedures used today were primarily optimized for identifying influenza viruses in poultry, not wild birds.

For example, virus isolation is the only available method that allows complete characterization and genetic sequencing. This procedure, typically done by inoculating wild-bird virus samples into chicken eggs, is time-consuming, expensive, requires special biosafety considerations and, importantly, might not work at all with some wild-bird viruses. Several laboratories are working on this problem, but resources are limited, and there is still an urgent need to develop and validate diagnostic tests specifically for wild birds.

The inconvenient truth is that our close relationship with domestic animals and wildlife puts us at increasing risk of an influenza pandemic, and, short of developing a vaccine, our best option is surveillance. The immediate pandemic threat being tossed our way is H5N1, but other viruses are lurking, and we had simply better be looking in the right direction, and working together, if we hope to catch the flu before it catches us. ■

Walter Boyce is director of the Wildlife Health Center, and co-director of the NIH Center for Rapid Influenza Surveillance and Research, at the School of Veterinary Medicine, University of California, Davis, California, USA.

1. http://www.fao.org/docs/eims/upload/231765/EW_Europe_aug07_ai.pdf
2. <http://www.gains.org/>
3. Gaidet, N. et al. *J. Wildlife Dis.* **43**, S22–S28 (2007).
4. <http://www.defra.gov.uk/animalh/diseases/monitoring/pdf/qra-recentdevelopments120707.pdf>

BOOKS & ARTS

The ultimate Romantic adventure

Napoleon's invasion of the Middle East enhanced rather than diminished the world's intellectual heritage.

Mirage: Napoleon's Scientists and the Unveiling of Egypt

By Nina Burleigh

Harper: 2007, 304 pp, \$25.95

Andrew Robinson

The unexpected invasion and occupation of Egypt in 1798–1801 by Napoleon Bonaparte's army, accompanied by 151 French scientists, scholars and artists, is a rich and shimmering subject. It was, as Nina Burleigh puts it, “the ultimate Romantic adventure”.

For more than three years, the French roamed the Nile valley and its surrounding desert, investigating every accessible detail of its archaeology and natural history — from pyramids and hieroglyphs, to crocodiles and scarab beetles. The wonders of ancient Egypt were revealed to Europe in the expedition's monumental publication, *Description de l'Égypte* (1809–28). Perhaps the most important discovery of all, in 1799, was the Rosetta Stone, which eventually allowed the pharaohs and their subjects to speak to the modern world.

Much has been written about this adventure from disparate angles. The historian of science Charles Coulston Gillispie chronicled its progress in a section of his magisterial *Science and Polity in France: The Revolutionary and Napoleonic Years* (2004). The British Museum Egyptologist Richard Parkinson discussed its significance to archaeology in *Cracking Codes: The Rosetta Stone and Decipherment* (1999). The literary critic Edward Said, in his influential *Orientalism* (1978), indicted Napoleon and his expedition for making scholarship subservient to imperialism.

Yet there has been no general account of the expedition's scientific and cultural aspects in English (Yves Laissus's colourful French book, *L'Égypte, une Aventure Savante* (1798–1801) has yet to be translated). This is the gap that Burleigh, an American journalist and author, aims to fill with *Mirage: Napoleon's Scientists and the Unveiling of Egypt*.

The challenge is formidable. A grasp of French, British and Middle Eastern history is essential; almost as important is an understanding of the sciences in 1800. Although disciplines were nowhere near as distinct as they are today, the expedition boasted specialists such as Joseph Fourier, Pierre Berthollet, Jules Savigny and Vivant Denon, in fields from mathematics and chemistry to natural history and art.

Indeed, one of the excitements of the expe-



More than 150 French scientists and artists catalogued the Nile valley's archaeology and natural history.

dition, as Burleigh points out, was its salon atmosphere. Among the experts relaxing in the palaces and gardens of Cairo, commandeered from the defeated Mamelukes, “architects debated with naturalists about animals and ancient structures, physicians and astronomers debated with the geographers about the meaning of the hieroglyphic script, the age of the ancient culture. These conversations among learned men manifested the highest ideals of the Enlightenment.”

Also challenging is to weave a clear and accurate narrative out of the fascinating but messy interactions of politics, scholarship and the military. Napoleon revered Newton and knowledge for knowledge's sake. But once he deserted Egypt in 1799 to grab power in Paris, relations between scientists and soldiers became tense and at times murderous. When the French general finally capitulated to the British in 1801, he wrote caustically to his opposite military number: “Several among our collection-makers wish to follow their seeds, minerals, birds, butterflies, or reptiles wherever you choose to ship their crates. I do not know if they wish to have themselves stuffed for the purpose, but I can assure you that if the idea should appeal to them, I shall not prevent them.”

Burleigh structures her book chronologically, more or less, yet tries simultaneously to focus on one field of endeavour per chapter. So we get: ‘The Inventor’ (Nicolas Conté, inventor of the graphite pencil), or ‘The Zoologist’

(Geoffroy Saint-Hilaire, whose work influenced Darwin's theory of evolution), or ‘The Stone’. The result is a somewhat confusing narrative in which, say, Napoleon's successor is first assassinated by a Muslim fanatic and then appears to oversee the despatch of field expeditions up the Nile. Although the individuals and their relationships come to life, and the descriptions of Egypt are vivid, these do not compensate for the lack of a coherent story.

Of the science, there is remarkably little; and some of that is misleading or wrong. A mere sentence is devoted to Gaspard Monge's theory of the mirage, despite the book's title. Devised from Napoleon's desperate desert march from Alexandria to fight the Battle of the Pyramids, in which soldiers were tormented by ‘water’ shimmering on the horizon, Monge's theory correctly accounts for mirages as the refraction of light by layers of air of differing density. Burleigh writes of “light and heat bouncing off the Earth's surface that created a mirror effect”.

The discussion of the Rosetta Stone is most unsatisfactory. It puts the hieroglyphs at the bottom, instead of at the top, of the stone. And it neglects several theories of phonetic elements in the hieroglyphs put forward before Jean-François Champollion's (by no means single-handed) phonetic deciphering of 1822. There is no mention, for example, of the phoneticism known to the medieval Arabs that Okasha el-Daly revealed in *Egyptology: The Missing Millennium* (2005). Elsewhere Burleigh states that

G. DAGLIORI/MUSÉE DU LOUVRE PARIS/THE ART ARCHIVE

Humphry Davy, rather than Benjamin Thompson, disproved the caloric theory of heat, and she seems unaware that William Thomson and Lord Kelvin are the same scientist.

In 2004, the Bibliotheca Alexandrina, the new library of the city where Napoleon landed, digitized and made available online and on CD the multi-volume *Description de l'Égypte*, originally issued by the French government. Meanwhile an empire-builder of the twenty-first century was wrecking the archaeological

sites and museums of another great Middle Eastern civilization.

Napoleon, for all his ruthlessness in Egypt, increased, rather than diminished, the world's intellectual heritage. The book that this extraordinary scientific adventure deserves has yet to be written. ■

Andrew Robinson is a visiting fellow of Wolfson College, Cambridge. His recent books include *The Last Man Who Knew Everything* and *The Story of Measurement* (Thames & Hudson).

Size matters, sometimes

Body Size: The Structure and Function of Aquatic Ecosystems

edited by A. Hildrew, D. Raffaelli and R. Edmonds-Brown

Cambridge University Press: 2007. 356 pp. £35, \$65

Steven Murawski

Shakespeare had it right when he had Pericles say, "Master, I marvel how the fishes live in the sea. Why, as men do a-land; the great ones eat up the little ones." Individual body sizes of organisms in aquatic ecosystems vary by a stunning 20 orders of magnitude — from free-living bacterial forms at 10^{-12} grams up to 180 tonnes for the blue whale, the largest animal ever to inhabit the planet. The distribution of animals of different sizes in a community dictates the structure and functioning of their ecosystem. For example, larger animals are much rarer than smaller ones, but they may still control the abundance of smaller plants and animals owing to factors such as their high rate of consumption and longer average lifespan.

On land, very large life forms, such as trees, can be found at the base of the food chain. Tiny organisms such as phytoplankton in the aquatic domain dominate primary production, where large predators prey on plants and animals that are a tenth to a thousandth of their own body size. Suspension feeding by straining particles suspended in the water (the aquatic equivalent of spider-web tactics) is commonplace. Parasitism, where predators are smaller than their prey, is comparatively infrequent.

Body Size: The Structure and Function of Aquatic Ecosystems focuses on size relationships between different components of aquatic ecosystems. Using this approach to interpret complex patterns provides the tools to help understand the consequences of human activity. For example, fisheries select the largest fish — with the result that top predators are harvested first. What are the implications of intensive and selective fishing for the stability and productivity of the aquatic ecosystem? One is that the increasing scarcity of larger fish may be associated with a concomitant rise in abundance of the smaller ones that are their usual prey.

Size-structured models help us interpret cascading impacts on other communities — for example, declines in zooplankton mean less grazing on phytoplankton, whose populations may therefore increase — and on a variety of other complex feedbacks in ecosystems. An important aspect explored in *Body Size* is the potential effect on nitrogen-cycling of removing larger organisms because they excrete more than smaller ones.

In the summary chapter — 'Body size in aquatic ecology: Important, but not the whole story' — the editors provide a frank assessment of where theory and observations regarding metabolic concepts in ecology align, and where they have yet to do so. For example, sev-

eral chapters question whether the relationship between body size and metabolic rate is fixed, and meta-analyses of field data point to some intriguing variations. Mid-water organisms on average have higher rates of metabolism than those of equivalent size in bottom-dwelling systems, perhaps because of their perpetual movement for predator avoidance. Freshwater invertebrate communities mostly support the fixed relationship, but those from a particular stream type that is replete with nutrients from woody debris, do not. We cannot yet say whether these anomalies are exceptions that prove the rule, or the result of one-off measurements or of an oversimplistic model of relationships governing disparate aquatic ecosystems.

Research on body size as an organizing principle in aquatic ecosystems, explored in this fine book, involves useful groupings of subsets of species and ecosystem types. The volume's worth is found in the series of chapters that are thematically organized around developing sub-areas of body size that allow more elaborate interpretation than a single unifying metabolic theory. This approach may need to be broadened if it is to be used to provide a systematic interpretation of the full variety of patterns observed in nature. ■

Steven A. Murawski is director of scientific programs and chief science adviser at National Marine Fisheries Service, NOAA, Silver Spring, Maryland 20910, USA.



Little and large: cleaner fish set to work on a green sea turtle, in decline because of fisheries' bycatch.

M. USHIO/IMAGEQUESTMARINE.COM



Did 'bubble' universes form during the Big Bang when a false vacuum sped up the expansion of matter?

The importance of nothing

The Void

by Frank Close

Oxford University Press: 2007. 176 pp.
\$19.50, £9.99

Lawrence Rudnick

In nothing, there is the potential for everything, from the blank canvas of the artist, to the creation myths of the world's religions, to the richness of the quantum vacuum. The contemplation of nothing can lead to paralysing circularity — is 'no dogs' the same as 'no cats'? It has also led science popularizer Frank Close to create *The Void* as a guidebook through nothing for the general reader.

Close begins with a charming section that relates the questions of his youth, such as the reality of the world before one's own birth. Here he waxes poetic, describing our physical bodies: "We have not been created out of nothing, but from a primeval 'ur-matter', atoms formed billions of years ago that have for a brief while been gathered into collections that think they are us." Sadly, such beautiful language is sparse in the rest of the book.

To the physical scientist, the vacuum is a fecund state; in some ways we parallel the struggles of the ancient Greeks who pondered the reality of space without matter. Although they flirted with the ideas of basic building blocks — atoms — between which empty space could exist, the aristotelian abhorrence of the vacuum dominated thinking for 2,000 years until the birth of experimental science. From his experiments on motion, Galileo inferred the pure motion of an object *in vacuo*, even though he couldn't produce one. The barom-

eter of Torricelli, the Magdeburg spheres of von Guericke, and the wine-and-water theatrics of Pascal led us to the mechanistic view of the vacuum: there was now a truly empty space whose properties could be measured — for example, the ability to transmit light but not sound.

Today's void is much more complicated, and Close guides the reader on the perilous journey past the battles over the existence of the ether, through Einstein's space-time and into the paradoxical world of the quantum vacuum. We sail the infinite sea of Dirac's antimatter states waiting to be excited into existence, to the Higgs vacuum from which all particles gain their mass, to the dark energy that accelerates the expansion of the Universe. And finally, to the origins of the Universe, where Close toys with Hawking's and Hartle's notions of imaginary time. It is an exhausting journey through the now-vital emptiness.

In the end, *The Void* is good for nothing — not great, but good. It covers very complicated concepts in a mostly accessible way, but lacks the graceful prose of either J. Barrow's *The Book of Nothing* or K. C. Cole's *The Hole in the Universe*. Often, Close struggles to make things understandable to the lay reader.

Cole, by contrast, simply reminds us of the alignment of iron filings around a magnet. For this journey, I would start with one of the earlier books, then try *The Void* for another perspective on the endlessly fascinating topic of nothing. ■

Lawrence Rudnick is distinguished teaching professor of astronomy at the University of Minnesota, Minneapolis, Minnesota 55455, USA.

EARTH OBSERVATION

A fresh perspective

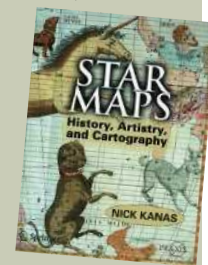
Earth From Space (A&C Black, 2007) by Andrew K. Johnston offers a view of the world from above, with more than 300 satellite photographs of Earth. Some are instantly identifiable as rivers, roads and cities; others, like fields in Garden City, Kansas, look like abstract art. Infrared shots of the Iraq-Iran border and radar images of Belgrade reveal the effects of war, environmental disaster and agriculture. The book covers the evolution of our attempts observe the Earth, from Wilbur Wright's plane-mounted camera to today's remote sensors. A chapter called 'Tools of the Trade' explains the technology used to take the photographs, leading in to chapters, such as 'Structure of the Land and Human Presence', that explore the forces that shape the world.



Nick Kanas's *Star Maps: History, Artistry, and Cartography* (Springer, 2007) adopts the

opposite perspective — looking up at the sky.

It chronicles attempts to chart the stars from ancient times to today. Alongside the familiar terrain of classical Western astronomy are star charts from China, Egypt and Mesopotamia. As well as explore the changing equipment in astronomy and cartography, the book covers the philosophies and personalities that saw star charts develop from images of gods and animals into the more scientific (although less beautiful) maps in use today.



Spies in the Sky (Praxis, 2007) by Pat Norris argues that reconnaissance satellites helped prevent outright conflict during the cold war era. It charts the development of satellite technology and the effect this had on international relations from the 1950s to the present day, before going on to speculate on how satellites might be used in future conflicts. The cold war sparked the demand for accurate, timely images of human activity across Earth that eventually led to the kind of equipment responsible for the book's stunning images.



Jennifer Meyer

HISTORY OF MEDICINE

Ageing of plastic surgery

According to legend, the fourth-century twin martyrs Cosmas and Damian, patron saints of physicians, successfully transplanted a leg from a black donor to a white recipient — a surgical triumph depicted 11 centuries later in this German painting. The discarded white limb on the floor is among the least graphic casualties in *A History of Plastic Surgery* by Paolo Santoni-Rugiu and Philip Sykes (Springer, 2007).

The volume's 400 or so figures are testament to three millennia of piecing body parts together and reconstructing or correcting them for functional or cosmetic purposes. The authors

have scrutinized rare ancient texts and drawings to gain insight into early medical practices, including the crucial adjuncts of anaesthesia, blood transfusion and treatment of infection, some of which survived for thousands of years.

The Greeks, the Romans, the Arabs, the Renaissance, and battlefields galore, contributed to advances in plastic surgery.

The authors' final figure is intended to give the reader "something to smile about". It charts abdominoplasty techniques from 1948 — depicting darkly comical stout stomachs and pendulous abdomens.



WÜRTTEMBERGISCHES LANDESMUSEUM, STUTTGART

The politics of mapping

Cartography and power have long gone hand in hand, as this pilgrimage plan illustrates.

Martin Kemp

Maps exercise a fascination — it is reassuring and even commanding to be able to locate ourselves and others from a kind of God's-eye perspective. To misquote Descartes, "I am mapped, therefore I am."

Cartographers chart the familiar and accessible for various purposes, from travel to property rights. They give us a vivid concept of far-off lands to which we shall never journey. Armies have used maps to prosecute wars, and worshippers follow them on pilgrimages, actual or imagined.

The Bible has many geographical passages, and nothing for certain can be learned about the text unless we first study them. The whole course of scripture is governed by regions, cities, deserts, mountains and seas.

A Christian with a grasp of biblical geography can be a surrogate eye witness to the momentous events central to his religion. At least, this is what Roger Bacon claimed in his *Opus Majus* of 1267, in a section dedicated to 'Mathematics in the service of theology'.

"If the reader understands the geographical operation of the four elements, hot and cold, dry and damp, and the effects of the mixtures of these four ... he will be able to grasp and delight in the pure and literal sense of the Scriptures, and be able to advance with pride and confidence to their spiritual meaning," he writes.

The earliest illustrated travel book was devoted to the Holy Land. In 1486, Bernard von Breydenbach, a canon at Mainz Cathedral in Germany, published an account of his pilgrimage in his *Peregrinatio in Terram*

Sanctum. He took with him the Dutch artist, Erhard Reuwich, who produced a great fold-out panorama centred grandly on the Dome of the Rock (Solomon's temple). von Breydenbach and Reuwich were more concerned with pictorial depiction and symbolic emphasis, however, than with precisely measured mapping.



The most notable of the early maps of the Holy Land based on systematic measurement resulted from the 1493 pilgrimage by Elector Friedrich III of Saxony. The six-sheet woodcut (pictured) was almost certainly produced by Lucas Cranach the Elder, who became Friedrich's 'ducal painter' in 1505. It is discussed in the third magisterial volume of *The History of Cartography, Cartography in the European Renaissance* (University of Chicago Press, 2007), edited by David Woodward. The Elector secured Cranach's income by granting him a monopoly of the sale of

medicines in Wittenberg and copyright in Bibles. Famed not least for his painted portrait of friend Martin Luther, Cranach was also a master print-maker. It is not known who provided the cartographic information.

Places with such resonant names as Galilee, Jerusalem, Bethlehem, Jericho and the River Jordan are duly assigned their places, and denoted by the little pictograms characteristic of Renaissance maps. The epic and perilous nature of Friedrich's pilgrimage is evoked by the galleons with billowing sails, the raging seas in the upper and middle left (the latter concealing a monstrous fish), and the excited landing party in their barge.

Here, as so often, religious devotion and territorial conflict are close companions. The Holy Land was the target of the crusading ambitions of popes and princes throughout the Middle Ages and the Renaissance. If we look at the ships crossing the sea, one is cresting proudly through the waves, bearing the Elector's shield on its sail. The lower one is a strange craft being blown forcefully in a contrary direction, and perhaps even travelling backwards. This seems to be Cranach's fanciful version of a hostile but thwarted Turkish man-at-war.

This piece of wishful thinking reminds us that cartography regularly served political ends. Cranach's map bears a militantly Christian message that contemporary viewers would not have missed. Looking at the map now, we realize that essentially the same contested interests of competing religions rule the unhappy history of the major sites today.

JEWISH NAT. & UNI. LIBRARY, JERUSALEM/ERAN LAOR CARTOGRAPHIC COLLECTION

Whole Earth comes into focus

To understand how our planet uses energy, we must integrate genetic data from microbial studies with satellite views of our planet.

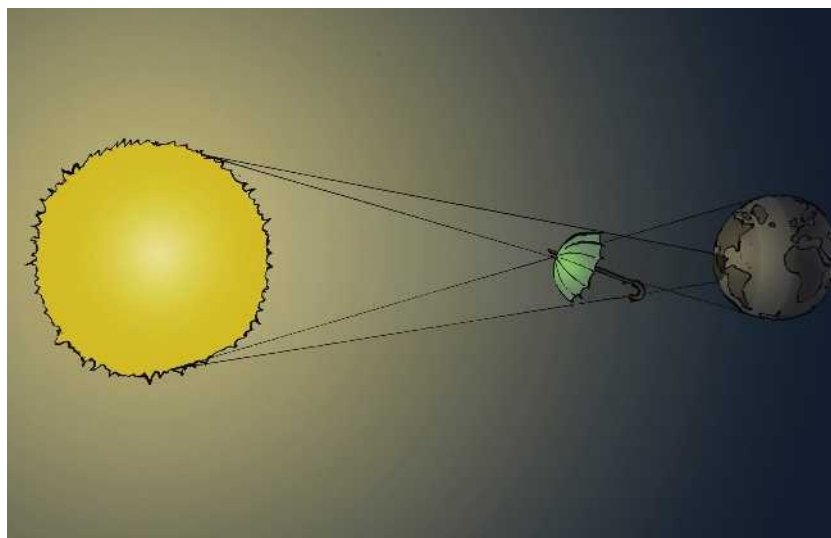
Stewart Brand

Two vastly different but complementary projects could transform our understanding of Earth. The long-standing mystery of how microbes run the world is closer to being solved, thanks to metagenomics — the DNA sequencing of whole populations of microbial life. And if a project to record fluctuations in the solar energy that reaches Earth gets back on track, we could begin to predict, and even manipulate, ecological changes on the planetary scale.

In 1966, I promoted the idea of photographing the 'whole Earth' from space, hoping that it would stimulate humanity's interest in its mega-habitat. The concept of the whole-Earth photo expanded profoundly in 1998 when the then US vice-president, Al Gore, envisaged a video camera in space permanently broadcasting a high-resolution real-time image of the sunlit side of our planet. The camera would reside at Lagrange-1, the point between the Sun and Earth where the gravitational pull is neutral. At this point in space, 1.5 million kilometres from Earth, an object orbits the Sun in synchrony with us. Gore's scheme was modified and eventually named the Deep Space Climate Observatory (DSCOVR). This project won approval from the US National Academy of Sciences and \$100 million from the US Congress, and the satellite was built.

DSCOVR's positioning, between the Sun and Earth, was to make it perfect for monitoring the planet's albedo, the amount of light Earth reflects. This information is crucial for improving climate models. The satellite was also to gauge Earth's total heat budget, measuring and guiding our progress in heading off global warming. It was also to provide much-needed calibration of the temperature readings from low Earth-orbiting satellites. And everyone with a computer screen would have been able to contemplate the bracingly real 'big here right now' image of our planet. Sadly, the satellite languishes in a warehouse outside Washington DC. The Bush administration mothballed the project in January 2001, and cancelled hopes for its launch in 2006.

Whether or not DSCOVR makes it to Lagrange-1, something similar surely must, and soon. We need to answer some crucial questions. Where exactly does solar energy go when it hits Earth? How much of it is absorbed by biological organisms? And how does biological life affect the physical energy flows in the atmosphere and oceans



that in turn drive climate dynamics?

When DSCOVR, or better, gets installed, the most important global activity it can monitor will take place at the scale of a bacterial gene. Metagenomics is giving us detailed access to the genes and gene communities of bacteria and archaea, 99% of which can't be cultured in the lab. For instance, analysis is revealing whole new metabolic pathways for oxygen and carbon production by the ocean's microbes.

These pathways must be understood, because their effects work on the planetary scale and could be harnessed as new energy sources. For example, the carbon tightly bound in lignin is broken down largely in the microbial soup of termite hindguts. Half of the world's oxygen is produced by ocean microbes, and they fix an unknown, but presumably enormous, amount of atmospheric carbon. Once ecologists open microbial black boxes such as the soil and gut, and when they understand the prolific transfer of genes between prokaryotes and viruses, we will begin to comprehend the extremely local interactions that take place within a biological neighbourhood. With the kind of close monitoring that metagenomics offers, predictability of ecology should improve.

This, perhaps, could be followed by engineering. To control global warming, we may one day resort to palliatives, such as a variable sunshade for Earth located at Lagrange-1 or increasing the activity of phytoplankton with deep-water nutrients or iron.

Both routes offer the ability to manipulate energy at the planetary scale, and each should be investigated incrementally. Solar

uptake by microbes can be probed locally with ocean experiments and then scaled up. And a sun-shield disk that can expand to a diameter of 2 kilometres would allow heat-budget changes to be studied with precision.

DSCOVR equipped with a variable sunshade could even act as a planetary thermostat, making fine adjustments in the amount of solar energy reaching Earth so that a stable climate regime is maintained. This vision requires much better understanding of temporal and regional energy budgets and their interplay with microbial life on Earth.

A unifying body of data, ideas, models and images of the whole-Earth system could inspire the public and may shift scientific thinking. In studying the energy dynamics of the Earth-Sun system while learning how our microbial partners manage to keep this planet comfortably terraformed for life, we would begin to step up to the full meaning of Earth stewardship. ■

Stewart Brand founded and edited the Whole Earth Catalog. He is president of the Long Now Foundation, co-founder of Global Business Network. His next book, *Whole Earth Discipline: An Ecopragmatist Manifesto*, will be published in late 2008 by Viking-Penguin.

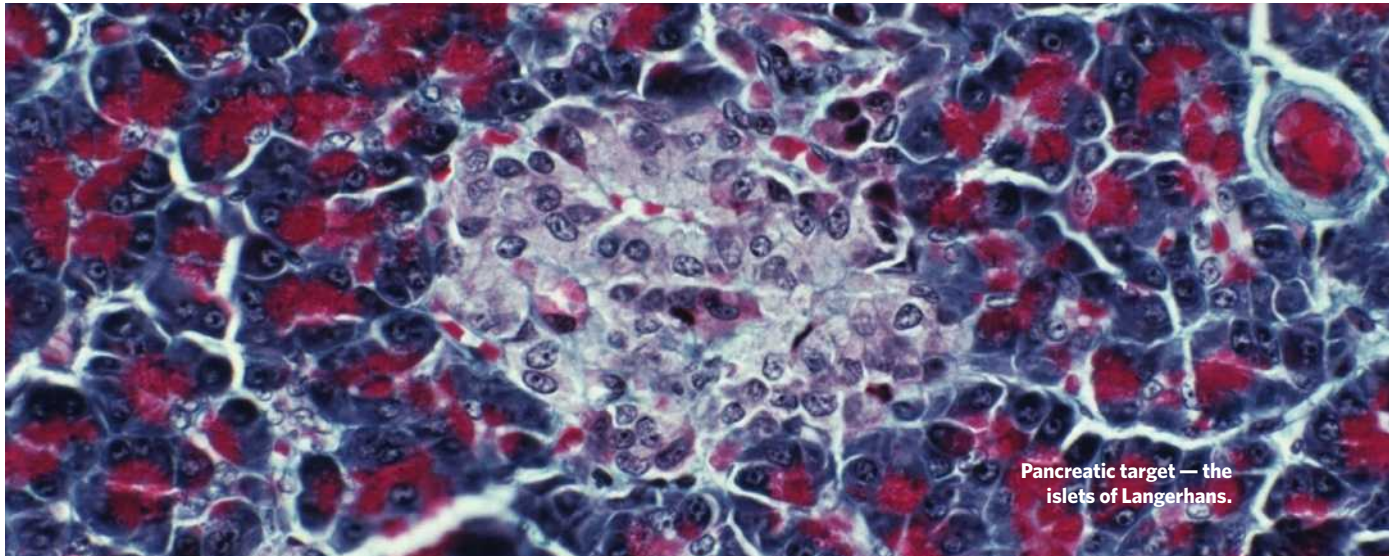
FURTHER READING

Handelsman, J. et al. *The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet* (Nat'l Acad. Sci., Washington DC, 2007); www.nap.edu/catalog.php?record_id=11902.
 NASA Deep Space Climate Observatory http://science.hq.nasa.gov/missions/satellite_53.htm
 Valero, F. *DSCOVR Mission Summary* http://cloud.ucsd.edu/dscovr/mission_summary.html
 Karl, D. M. *Nature* **415**, 590–591 (2002).
 Warnecke, F. et al. *Nature* **450**, 560–565 (2007).

J. KAPUSTA

ESSAY

NEWS & VIEWS



PHOTOLIBRARY.COM

DIABETES

Missing links

Bart O. Roep

Researchers have now probably pinpointed all the genes in the MHC genomic region that are risk factors in type 1 diabetes. As the MHC is unusually rich in genes involved in immunity, this is truly exciting.

If there were 'Guinness World Records' for genomic regions, the major histocompatibility complex (MHC) would have several entries. It is the most gene-dense region in mammals, and variations in the human MHC are linked to susceptibility to more common diseases than any other region in the human genome. Moreover, MHC-associated diseases include almost all autoimmune diseases. But because of high gene density in the MHC region and the localization of these genes very close to each other, determining which gene, or group of genes, is associated with a specific disease is akin to looking for a needle in a haystack. For example, some variants of the MHC-localized genes that encode HLA proteins are known to determine different susceptibilities to the autoimmune disorder type 1 diabetes — ranging from increased vulnerability to increased protection. But this seems not to be the whole story, and the identities of other HLA genes suspected to be associated with type 1 diabetes remain elusive. Nejentsev *et al.*¹ (page 887 of this issue) now zoom in on the MHC to identify these missing links.

In individuals genetically susceptible to type 1 diabetes², misregulation of the immune system results in immunological intolerance towards insulin-producing β -cells, which are part of cell populations in the pancreas called

the islets of Langerhans. Type 1 diabetes is characterized by the inflammation of islets of Langerhans and, subsequently, the selective destruction of the β -cells. Consequently, patients become permanently dependent on insulin injections.

The highly variable HLA genes are clustered in two loci, or positions, in the MHC region on chromosome 6, separated by one genomic map unit (1 centimorgan). They encode two classes of proteins: the HLA class I and the HLA class II molecules. These proteins are expressed on cell surfaces, where they present both self and foreign peptide antigens to different types of T cells.

Which MHC-localized genes are risk factors for autoimmune diseases is extensively debated. For example, although gene variants (alleles) of HLA class II genes are known to impose the greatest predisposition to type 1 diabetes, the MHC region has long been suspected of harbouring other genes contributing to the risk of this disease³. Many MHC genes have immunological functions, however, making them suspects in contributing to autoimmune responses, and the unusually high density of genes in the MHC offers an almost endless list of potential candidates⁴. What is more, because of their strong genetic linkage (linkage disequilibrium), most of the MHC

genes are inherited together⁴, which makes it difficult to separate their effects. Thus, several attempts at pinpointing additional culprits in the pathology of type 1 diabetes have been unsuccessful. Consequently, the entire MHC region has been deemed a 'super-locus', consisting of a cornucopia of discrete risk loci for type 1 diabetes⁵.

In the latest attempt at disentangling risk factors for type 1 diabetes, Nejentsev *et al.*¹ have mapped all remaining MHC genes known to be associated with an increased risk of the disease to several alleles of the HLA class I *HLA-A* and *HLA-B* genes. This provides evidence that HLA class I genes (as well as, and independently of, HLA class II genes) make a strong contribution to the pathology of this disease. Although the authors leave room for other minor MHC contributors, their work suggests that classic antigen-presenting HLA class I and II molecules, and nothing else, account for MHC-associated genetic susceptibility to this disease. All other possible associations are probably attributable to their strong genetic linkage with HLA class I and II genes.

The association of HLA class I genes with type 1 diabetes has been reported previously⁶. But Nejentsev and colleagues' findings stand out from earlier observations in two respects. First, their data are based on an experimental

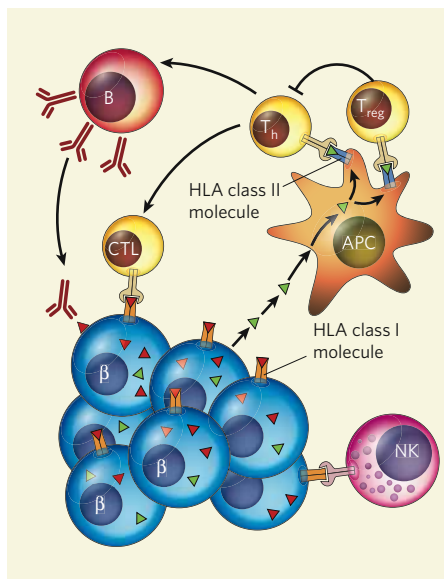


Figure 1 | Type 1 diabetes and the contribution of HLA class I and class II proteins. In the pathogenesis of type 1 diabetes, HLA class II molecules on the surface of antigen-presenting cells (APCs) present fragments of pancreatic β -cell proteins to a subclass of T cells called T helper (T_h) cells. In turn, the autoreactive T_h cells stimulate cytotoxic T cells (CTLs) to attack insulin-producing β -cells (β); HLA class I proteins on the β -cell surface present β -cell peptides to CTLs. T_h cells also stimulate B cells to produce antibodies against cells of the islets of Langerhans. Moreover, KIR receptors on natural killer cells (NK) can also engage with HLA class I proteins on the islet cells and, depending on the type of KIR, help to activate or inhibit autoimmune responses against β -cells. Normally, regulatory T cells (T_{reg}), which are activated by β -cell peptides presented by HLA class II proteins on APCs, inhibit autoimmune T_h -cell function, thereby suppressing disease.

approach that is sufficiently powerful to overcome the confounding effect of HLA class II genes — they used a large number of patients and control subjects and performed a comprehensive scan of the entire MHC region.

Second, to rule out the influence of HLA class II genes, the authors used the statistical method of recursive partitioning and regression, which avoids unwarranted assumptions, thereby forming a tree of known, disease-risk-associated HLA class II genes in which secondary associations were tested. They then meticulously validated the contribution of these HLA class II genes to confirm that no residual effects, other than the detected HLA class I contribution, was responsible for the observed associations.

Why do certain HLA genes affect disease susceptibility? Although evidence for a link between the disease association of these genes and the function of their protein products is scarce, it is tempting to speculate on such a link. Variation in the HLA genes leads to a large variety of peptide-binding molecules that generate specific immune responses. For

example, HLA proteins play an essential part in shaping the developing T-cell repertoire in the thymus, limiting autoreactivity and guiding antigen-specific T-cell responses.

This feature might explain why certain alleles of HLA class I and II genes confer an increased risk of type 1 diabetes. Both pro-inflammatory, autoreactive T helper cells and cytotoxic T cells can recognize β -cell-derived peptides presented by class II and class I HLA proteins, respectively^{7,8} (Fig. 1). These responses are reflected in the inflammation of islets of Langerhans in type 1 diabetes, and in the β -cell destruction that occurs after transplantation of islets into patients with diabetes. Moreover, HLA class I proteins bind to KIR receptors on natural killer cells. In response, these cells of the innate, or nonspecific, immune system contribute to the regulation of adaptive (specific) immune responses. Thus, HLA class I proteins link innate and adaptive immunity, and their misregulation in disease can lead to attack on the individual's own β -cells.

The innovative analytical approach of Nejentsev *et al.*¹ underscores the outstanding importance of HLA class I genes not just in the pathogenesis of type 1 diabetes but, conceivably, in other HLA-associated autoimmune diseases. It also suggests that, with sufficiently powerful experimental tools and appropriate

study design, specific HLA genes can be implicated in specific autoimmune diseases. MHC genes other than HLA class II genes have also been linked to other autoimmune diseases such as rheumatoid arthritis, coeliac disease, systemic lupus erythematosus and myasthenia gravis. So it is crucial to define whether these associations can be attributed to class I alleles and to elucidate the involvement of this class of HLA molecules in the general mechanisms of autoimmune diseases. For type 1 diabetes, it is now beyond doubt that the contribution of HLA to the process of β -cell autoimmune destruction is more than just guilt by genetic association. ■

Bart O. Roep is in the Department of Immunohaematology and Blood Transfusion, Leiden University Medical Center, Albinusdreef 2, NL-2333 ZA Leiden, the Netherlands. e-mail: boroep@lumc.nl

1. Nejentsev, S. *et al.* *Nature* **450**, 887–892 (2007).
2. Atkinson, M. A. & Eisenbarth, G. S. *Lancet* **358**, 221–229 (2001).
3. Aly, T. A. *et al.* *Proc. Natl Acad. Sci. USA* **103**, 14074–14079 (2006).
4. Horton, R. *et al.* *Nature Rev. Genet.* **5**, 889–899 (2004).
5. Zavattari, P. *et al.* *Hum. Mol. Genet.* **10**, 881–889 (2001).
6. Nakanishi, K. *et al.* *Diabetes* **42**, 1086–1093 (1993).
7. Roep, B. O., Arden, S. D., de Vries, R. R. & Hutton, J. C. *Nature* **345**, 632–634 (1990).
8. Pinkse, G. G. *et al.* *Proc. Natl Acad. Sci. USA* **102**, 18425–18430 (2005).

CONDENSED-MATTER PHYSICS

Shear madness

Alan T. Dorsey and David A. Huse

Cooled to temperatures just above absolute zero, solid helium starts to behave very oddly. But its 'supersolid' behaviour might just be the result of imperfections that change the bulk properties of the crystal.

In 2004, Eun-Seong Kim and Moses Chan placed a sample of solid helium, ^4He , in a torsional oscillator at a temperature of about 0.1 kelvin, and allowed it to twist a little. What they observed^{1,2} ensured that the properties of solid helium would become a hot topic. A portion of the solid seemed to detach itself and flow through the rest without encountering any frictional resistance — a 'supersolid' seemed to have formed.

On page 853 of this issue³, Day and Beamish report changes in a material property known as the shear modulus — which measures a solid's rigidity, or resistance to elastic deformation — in solid helium. The changes closely mirror the results of the torsional-oscillator experiments. Two possibilities immediately suggest themselves: first, that the onset of supersolidity might strongly affect solid helium's elastic properties; or second, and rather more prosaically, that changes in the solid's elasticity might have mimicked the effects of supersolid behaviour in the previous measurements.

Helium is the noblest of elements: the interactions between even its own atoms are so weak that it solidifies only under intense pressure. If this pressure is reduced to below about 25 atmospheres at absolute zero, the quantum-mechanical fluctuations of the atoms' positions become so large that the solid melts, becoming a 'quantum liquid'. No crystalline solid is perfect — there are always some vacancies in the crystal lattice where atoms are missing — and in 1969 Alexander Andreev and Ilya Lifshitz⁴ proposed that helium's large quantum fluctuations might, at zero temperature, stabilize a dilute gas of vacancies within the solid. Atoms of the prevalent isotope ^4He are bosons (they have zero spin), and so vacancies in solid ^4He can also be thought of as bosons. The vacancies can thus condense to form an exotic phase known as a Bose–Einstein condensate that suffuses the solid. This 'supersolid' phase would share some properties with a superfluid — namely, frictionless flow — but at the same time have a non-zero shear

modulus, a defining characteristic of a solid.

Until Kim and Chan, however, no one had had much luck in finding experimental evidence for supersolidity. Their torsional oscillator was a rather simple piece of apparatus. They attached a 'bob' containing the helium sample to the end of a torsion rod. A torque applied to the bob caused it to twist and rotate back and forth, with an oscillation period set by the ratio of the bob's moment of inertia (a measure of the amount and distribution of mass being rotated) to the torsional stiffness of the rod. If some of the material in the sample stops participating in this rotation, the moment of inertia of the bob decreases, and the period drops. This is precisely what happens with liquid ^4He , which is a well-known superfluid: as the temperature is reduced below that at which the superfluid forms, the superfluid component stops rotating and the oscillator's period drops. Kim and Chan's same result^{1,2} with solid ^4He has now been reproduced in at least four other laboratories.

A supersolid can exhibit other anomalies, for instance in the speed at which sound passes through it. Sound speed depends on the shear modulus of the solid, as well as the density of the superfluid component. To assess why the solid behaves in the way it does, it is thus important to measure the shear modulus independently of the superfluid density. This is precisely what Day and Beamish have now done with solid helium.

Again, the authors' experiment³ is conceptually simple. They placed solid ^4He between two parallel plates, known as piezoelectric shear transducers. They moved one plate, the driving transducer, in a direction parallel to the second plate. The solid helium transmits the resulting elastic shear stress between the plates, and this is measured by the second transducer. Day and Beamish find³ that the shear modulus of helium rises by up to 10% as the temperature is reduced from 0.2 to 0.02 kelvin. More significantly, the temperature dependence of this large increase in shear modulus closely tracks the changes in period in the torsional-oscillator experiments.

When these results were first presented at a small workshop on the supersolid state of matter in Minnesota in late July 2007, the reaction of some of us in the audience was that the large change in the shear modulus might be the primary phenomenon, and the much smaller change in the period of the torsional oscillator a related side effect. The various torsional-oscillator experiments are now being re-examined, to estimate the shear stresses and strains that occur within and between the various parts of the assemblies, and to ask if changes in shear stiffness of the helium might cause the observed changes in the oscillation period⁵.

So what might be the root cause of the observed effects? It has long been known that a solid's strength and shear modulus are not set solely by the intrinsic nature of a perfect

crystalline solid, but also depend strongly on defects such as dislocations and grain boundaries. The same seems to be true for the torsional-oscillator experiments: the results depend on the quality of the crystal, with the largest effects seen in the most defective samples.

Day and Beamish³ explain their results very plausibly as the behaviour of dislocations and trace ^3He impurities in the solid ^4He : when dislocations can move in response to a shear stress, they relax the stress and so lower the shear modulus. Impurities of ^3He are thought to bind weakly to dislocations in solid ^4He at low temperatures and so restrict their motion. As the temperature is increased, the trace ^3He atoms unbind from the dislocations, allowing the dislocations to move more freely. The observed dependence of the shear modulus on the concentration of ^3He is consistent with this idea. This is not to say that supersolidity must have no role: careful simulations⁶ and modelling⁷ find superfluid flow along dislocation lines in solid ^4He .

It seems clear now that both the earlier

torsional-oscillator experiments^{1,2} and Day and Beamish's new experiment³ are probing the quantum-mechanical behaviour of structural defects in imperfect helium crystals. But a thorough understanding of the measured effects remains elusive. Are we studying supersolidity, the quantum properties of dislocations, both — or perhaps something else entirely? ■

Alan T. Dorsey is in the Department of Physics, University of Florida, Gainesville, Florida 32611-8440, USA. David A. Huse is in the Department of Physics, Princeton University, Princeton, New Jersey 08544, USA.

e-mails: dorsey@phys.ufl.edu;
huse@princeton.edu

1. Kim, E. & Chan, M. H. W. *Nature* **427**, 225–227 (2004).
2. Kim, E. & Chan, M. H. W. *Science* **305**, 1941–1944 (2004).
3. Day, J. & Beamish, J. *Nature* **450**, 853–856 (2007).
4. Andreev, A. F. & Lifshitz, I. M. *Sov. Phys. JETP* **29**, 1107–1113 (1969).
5. Clark, A. C. & Chan, M. H. W. preprint at www.arxiv.org/abs/0711.3619 (2007).
6. Boninsegni, M. et al. *Phys. Rev. Lett.* **99**, 035301 (2007).
7. Toner, J. preprint at www.arxiv.org/abs/0707.3842 (2007).

RELATIVITY

Still special

Giovanni Amelino-Camelia

Is special relativity a clapped-out classical theory, to be replaced by a shiny new quantum model as soon as possible? On the contrary, it would seem: the theory still has a youthful ability to surprise us.

Since Albert Einstein introduced it to the world in 1905, the special theory of relativity has embodied the journey of modern physics from an 'intuitive' description of the world to a deeper level of understanding — an understanding at first profoundly baffling to established ways of thinking. Concepts such as the equivalence of mass and energy, embodied by the formula $E = mc^2$; the existence of an unbreakable speed barrier, the speed of light *in vacuo*, c ; and the paradox of two twins who, by dint of experiencing different accelerations through space, can age by different amounts, have all stamped themselves on the public's consciousness (Fig. 1, overleaf). At the same time, special relativity has provided a reliable description for an ever-growing list of physical phenomena.

Writing in *Physical Review Letters*, Cubero et al.¹ add to that list, establishing how special relativity affects certain equilibrium properties of a gas of idealized particles. The work is symbolic of special relativity's odd position in the pantheon of modern physics theories: continuously tested and always successful; yet also disparaged as not really 'belonging' in the grander scheme of things. To physicists, special relativity was only really ever a fundamental theory for 11 years: in 1916, it ceded that title

to Einstein's general theory of relativity. General relativity incorporates special relativity's maximum-speed principle into a comprehensive theory of gravitational phenomena, and through that arrives at a description of gravitational acceleration as a consequence of the curvature of space-time.

Special relativity is thus a humble 'effective theory', valid only as an approximation under certain conditions — specifically, that gravitational effects such as the curvature of space-time can be ignored. But, as luck would have it, physicists have devoted most of the century since Einstein first dreamt up his relativity to studying the quantum properties of particles in just such cases. The particle physicist at a laboratory such as CERN does not need to worry about the local topology of space-time when smashing particles together: particle physics' ultra-successful 'standard model' is built on the special, not the general, theory of relativity.

Since the end of the 1990s, however, the blanket application of special relativity in these instances has come under renewed attack. The reason for this is that gravity, in its general-relativistic description, is the only one of the four fundamental forces of nature that is still described using the rules of classical mechanics. Pursuing the hypothesis of a 'quantum

gravity', one encounters the possibility of small, but non-negligible quantum-curvature effects, even where the average curvature is zero.

This could have practical implications for, say, the observation of objects in the distant Universe. The particles that we observe from so far away (mainly photons of light) propagate for billions of years, and quantum-curvature effects, although individually tiny, could accumulate to an appreciable effect that might, for example, affect our estimates of the objects' distance. But in all the astrophysical tests so far performed to test this hypothesis, special relativity comes out tops, with no evidence of deviations from its predictions. A more definite picture will emerge only after other crucial tests are performed, such as γ -ray observations soon to be conducted with NASA's GLAST space telescope. But at present even I — a researcher deeply involved in all the quantum-curvature speculation — must admit that special relativity seems in remarkably rude health.

In fact, rather than discovering regimes to which special relativity does not apply, we are actually learning how to use it to describe even more areas of physics. Cubero and colleagues' analysis¹ of the special-relativistic equilibrium properties of a gas of particles is a case in point. Although Einstein's original formulation provided straightforward prescriptions for attributing special-relativistic properties to each particle in such a multi-particle system, it has not always been easy (and has sometimes been terribly hard) to derive from these properties a macroscopic, statistically valid description of the system.

The authors find a clever and reliable way to simulate numerically a dilute, one-dimensional gas consisting of two species of particle, and study macroscopic properties such as temperature and the velocity distribution of the particles. The statistical distribution of the particle velocities in the gas that emerges clearly favours a description proposed² on the basis of a maximum-entropy principle that has been criticized for lacking explicit compatibility with relativity. The authors also succeed in introducing a concept of 'proper' temperature such that, in agreement with the principles of special relativity, observers moving differently with respect to the gas will agree on the temperature's value.

As so often happens in the development of young theories — at 100 years old, special relativity is still looking remarkably youthful — these results clarify some issues, but also present new challenges. In particular, the authors' analysis must be generalized from one to three spatial dimensions, the number we really care about. This will require getting rid of the simplification introduced by the authors that the point particles in the gas interact only when they actually touch. In one dimension, touching collisions are rather frequent, and this approximation is appropriate; but in three dimensions, contact collisions would not be frequent enough to allow the establishment of

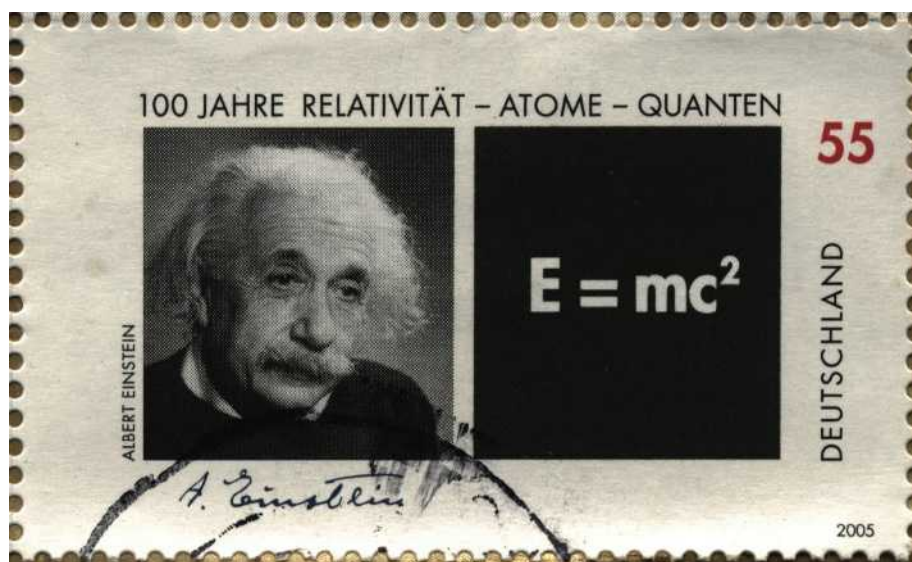


Figure 1 | Einstein equals most celebrated, squared. The equation embodying mass–energy equivalence has become, along with its creator, the most recognizable symbol of modern physics — here on a German stamp to commemorate the 2005 'Year of Physics', the 100th birthday of special relativity.

a definite temperature. For a complete description, interactions at a distance must also be considered.

Such caveats do not detract from the success, exemplified by Cubero and colleagues' work¹, of special relativity in characterizing the properties of complex systems. Even those of us speculating about a quantum version of the theory, and at present concentrating on very simple systems, might do well to take a lesson

from that — this theory's retirement might be some time coming yet.

Giovanni Amelino-Camelia is in the Department of Physics, Università di Roma La Sapienza, and the Sezione Roma 1, INFN, Piazzale Moro 2, Rome 00185, Italy.

e-mail: giovanni.amelino-camelia@roma1.infn.it

1. Cubero, D., Casado-Pascual, J., Dunkel, J., Talkner, P. & Hänggi, P. *Phys. Rev. Lett.* **99**, 170601 (2007).
2. Jüttner, F. *Ann. Phys.* **34**, 856–882 (1911).

CANCER

Immune pact with the enemy

Cornelis J. M. Melief

Progress comes from the latest investigations into a long-standing question in immunology — the role of the immune system in maintaining small, potentially cancerous lesions in a state of dormancy.

The current belief about treating cancer is that tumour cells need to be eradicated as quickly as possible, so as to halt tumour growth and spread, and to prevent or delay the death of the patient. The startling results of Koebel *et al.*¹ (page 903 of this issue) demonstrate that considering cancer as a fatal disease is not always appropriate*.

The authors show in mice that induction of cancer by methylcholanthrene (MCA), a tar component of the kind found in cigarette smoke, causes an initial wave of deadly tumours affecting many animals, after which the surviving mice show no evidence of growing tumours. Deceptively, however, dormant tumours still exist in these apparently healthy mice, and these are kept in check by

*This News & Views article and the paper concerned¹ were published online on 18 November 2007.

the immune system. This state can be fatally disrupted by suppression of the immune system, allowing the dormant tumours to wrest themselves from immune control and grow, spread and kill their host.

The concept of tumour dormancy, or latency, has a long history, and is well documented in animal studies and in anecdotal clinical observations^{2–5}. Also, it has long been suspected that immunosuppression can activate dormant tumours. For example, a cancer — malignant melanoma — developed in the recipient of a kidney transplant from a donor who had been treated for melanoma 16 years earlier, and was considered cured⁵. The cancer in the recipient was found to be of donor origin: apparently, the immunosuppression required to prevent the recipient's rejection of the kidney triggered regrowth of the cancer



50 YEARS AGO

"What mean these stones?" — The Bible is a library of books, not all of the same literary or historical importance. In some cases (especially in the Old Testament) the books themselves are composite and contain ancient fragments embedded in a more recent framework. When, therefore, Prof. Burrows suggests that archaeological researches in Palestine back up statements appearing in the writings, it must be clearly understood that in no sense can the verbal accuracy of the whole be demonstrated; all that can be said is that many topographical descriptions can be shown to be reasonably correct.

From *Nature* 7 December 1957.

100 YEARS AGO

In proposing the toast of "The Royal Society" at the anniversary dinner on Saturday last, Lord Dunedin referred to the popularisation of science as one of the functions of a society which exists for the promotion of natural knowledge. This remark provides the subject of a letter by an anonymous correspondent in Tuesday's *Times*. The writer urges that the neglect of science in this country is largely due to the indifference shown by scientific men to the intellectual interests of the average reader. Few men of science make any attempt to describe their investigations in language which can be understood by men of culture without special scientific knowledge, and it is scarcely too much to say that most investigators are so closely absorbed in their particular researches that whether the world in general knows anything of the results or not is regarded as no concern of theirs. This spirit, and the obscure and diffuse manner in which scientific investigations are often described, are to be deplored.

From *Nature* 5 December 1907.

that had been kept under immune control in the donor for all those years. In none of these previous reports, however, was the dormant state actually visualized, and the component of the immune system involved in maintaining dormancy was not identified.

Koebel *et al.*¹ take strides forward in both respects. After the initial wave of MCA-induced deadly tumours, dormant lesions in immunocompetent mice — those with a fully functional immune system — developed into progressive tumours only after treatment that resulted in the depletion of immune-system cells known as T lymphocytes (T cells), or neutralization of the cytokines interleukin-12 or interferon- γ , which are involved in adaptive immunity. Depletion of cells called natural killer cells, which are more broadly acting but less specific immune agents, had no effect. These results point to highly specific, adaptive T-cell immunity as the component of the immune system that maintains dormancy. Interestingly, in a different tumour model, immunization with tumour cells can generate antibodies that also contribute to dormancy⁴.

Koebel *et al.* also found that many cells in the stable, dormant lesions showed morphological features reminiscent of those in progressively growing, MCA-induced cancers. Like the growing cancers, the stable lesions were infiltrated by immune cells, including T cells, indicating that they were immunogenic (that is, they were being recognized by the immune system). But there was a much lower percentage of proliferating cells and an increased incidence of cell death. Transient culture of cells from dormant lesions yielded atypical fibroblast-like cells that grew out as tumours when injected into immunodeficient, but not immunocompetent, mice. Even in some immunocompetent animals, however, stable lesions occasionally escaped from dormancy and became cancerous. But these lesions could do so only if they had lost their immunogenicity, as indicated by their subsequent ability to grow in immunocompetent host animals.

Thus, Koebel and colleagues' work for the first time characterizes a state of tumour dormancy. The hallmarks of this state are stable lesions of transformed immunogenic cells, which are controlled by the host's adaptive immune system in a condition dubbed 'equilibrium' because of its dynamic nature. Obviously, this is a precarious situation — loss of either immunocompetence or immunogenicity can lead to tumour outgrowth, as the authors show.

The implications of this work are far-reaching. First and foremost, the description and visualization of dormant lesions offers an opportunity to characterize their molecular signatures, as determined by their gene-expression profiles, and to compare these signatures with those of the lesions that became cancerous even in immunocompetent hosts. Indeed, such understanding may lead to the development of new treatments, including non-immune-drug

interventions, to turn overt cancers into less aggressive, stable lesions. Second, as Koebel *et al.*¹ point out, cancer immunotherapy can aim not only at complete tumour eradication, but also at establishing tumour equilibrium by encouraging the production of interferon- γ -producing, tumour-specific T cells. Indeed, spontaneous T-cell infiltration into human cancers is now increasingly recognized as a favourable prognostic sign, independently of other indicators^{6–8}.

Third, this model of dormancy has striking parallels with the chronic infection caused by *Mycobacterium tuberculosis*, one of the world's most successful pathogens. Typically, an asymptomatic or latent infection is established, which can last for decades before the pathogen is reactivated and clinical tuberculosis ensues. Such processes often coincide with a phase of immune suppression^{9,10}. Indeed, *M. tuberculosis* is thought to use special bacterial gene products to maintain latency, and it is tempting to speculate that dormant tumours use similar tricks to avoid being eradicated.

Fourth, a more intense search for dormant tumours is warranted — particularly for those tumours induced by chemicals, such as may be present in cigarette smokers, given that Koebel and colleagues' mouse system mimics that situation. Obviously, dormant tumours in smokers would pose a threat, because they can awaken and become overt cancer, at which time it is usually too late for effective therapy¹¹. It may be difficult to detect truly dormant lung cancers. But perhaps patients with breast cancer offer another route for investigation: up to 22 years after undergoing a mastectomy, one-third of patients reportedly have evidence of circulating 'tumour' cells without any evidence of disease³. Are these cells also kept in check by immune responses, or are they controlled by other mechanisms?

A final, unwelcome, thought prompted by the new results concerns the treatment of cancer patients with immunosuppressive chemotherapy or irradiation. A downside of such treatment could be the escape of dormant tumour cells from immune control. Dormant cells themselves are likely to be less susceptible to these treatments, which primarily target rapidly dividing cells.

Cornelis J. M. Melief is in the Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Albinusdreef 2, PO Box 9600, 2300 RC Leiden, the Netherlands. e-mail: c.melief@lumc.nl

1. Koebel, C. M. *et al.* *Nature* **450**, 903–907 (2007).
2. Allan, A. L. *et al.* *Breast Dis.* **26**, 87–98 (2007).
3. Vessella, R. L. *et al.* *Cancer Biol. Ther.* **6**, 1–9 (2007).
4. Uhr, J. W. & Marches, R. *Semin. Cancer Biol.* **11**, 277–283 (2001).
5. Mackie, R. M. *et al.* *N. Engl. J. Med.* **348**, 567–568 (2003).
6. Galon, J. *et al.* *Science* **313**, 1960–1964 (2006).
7. Sato, E. *et al.* *Proc. Natl Acad. Sci. USA* **102**, 18538–18543 (2005).
8. Piersma, S. J. *et al.* *Cancer Res.* **67**, 354–361 (2007).
9. Andersen, P. *Trends Microbiol.* **15**, 7–13 (2007).
10. Leyten, E. M. *et al.* *Microbes Infect.* **8**, 2052–2060 (2006).
11. Bianchi, F. J. *Clin. Invest.* **117**, 3436–3444 (2007).

MICROBIOLOGY

Bilingual bacteria

Matthew R. Parsek

Many bacteria use chemical signals to coordinate group behaviour. A signal that suppresses virulence has been identified in the bacterium that causes cholera, and could be a new therapeutic target.

Like people, many bacteria do things in groups that they don't do on their own. These communal activities can be spectacular; the marine bacterium *Vibrio fischeri*, for example, produces bioluminescence in the light organs of deep-sea fish. But bacterial group behaviour can also be deadly — many bacteria become virulent only when they reach a certain local concentration. Such coordinated actions require bacteria to 'talk' to each other by sending chemical signals, a process known as quorum sensing. Although some of the molecules involved are known, it is likely that there are many more.

It's been known for a few years that *Vibrio cholerae*, the bacterium that causes cholera in humans, is 'bilingual' — that is, it uses two distinct signalling molecules¹ to suppress its virulence². One of these signals is a molecule that many species of bacteria use for quorum sensing³, but the identity of the second signal has remained a mystery. In this issue, Higgins *et al.*⁴ report the structure of this second signal. Their discovery represents a new structural class of quorum-sensing signal that may be exclusive to *Vibrio* bacteria, making it a possible lead for drug discovery.

In general, quorum sensing is straightforward: bacteria release signals into the surrounding environment; if the signals reach a critical concentration, they are detected by bacteria in the vicinity and this stimulates a response. In *V. cholerae*, quorum sensing proceeds through two parallel systems¹, either of which is sufficient to independently initiate a response. The first of these involves the AI-2 molecule (Fig. 1), a signal used by many species of bacteria. AI-2 is detected by the sensory proteins LuxP and LuxQ, which are associated with the bacterium's cell membrane.

Although the identity of the signal in the second system was unknown, the enzyme responsible for producing the signal had been identified as the CqsA protein. The second signal is thought to be detected by a putative membrane-associated sensor called CqsS. Both systems in *V. cholerae* funnel information into the same signalling cascade (through the transducing proteins LuxU and LuxO) so that an analogous functional response is produced in each case: LuxO is inactivated, resulting in increased activity of HapR, a negative regulator that represses expression of virulence genes.

In their detective work identifying the unknown quorum-sensing signal of *V. cholerae*, Higgins *et al.*⁴ took their cue from the known biochemistry. They introduced the *cqsA* gene

into *Escherichia coli*, creating a recombinant strain that produces much more signal than the parent *V. cholerae* species. By extracting the culture fluids of the *E. coli*, the authors obtained a mixture of compounds that they separated into its constituent parts. They next tested the purified compounds on a strain of *V. cholerae* that had been engineered to emit light in response to the unknown signal. Certain fractions of the mixture were 10,000 times more active than controls.

Using a combination of spectroscopic techniques, Higgins *et al.* then identified the active compound as 3-hydroxytridecan-4-one (Fig. 1). They confirmed this by preparing the compound chemically, and testing the synthetic version in their activity assay. The authors were finally able to obtain sufficient material from cultures of natural *V. cholerae* for analysis, and so to prove conclusively that 3-hydroxytridecan-4-one is the second signal for this bacterium.

So why does *V. cholerae* adopt a 'belt and braces' approach to quorum sensing, using two parallel systems when one would be sufficient? There are many potential explanations, one of which relates to the distinct chemical properties of the two signals. These properties might influence the stability or rates of diffusion of the signals, perhaps making one molecule superior to the other for quorum sensing in a particular environment. Higgins *et al.*⁴ suggest that, because *V. cholerae* encounters environments that are rich in other AI-2-producing bacteria (such as the large intestine), the AI-2 system might be used for interspecies signalling. Conversely, they propose that the specificity of the Cqs signal for the *Vibrio* genus makes it ideal for quorum sensing with other bacteria of the same species. To test this hypothesis, it will be necessary to show that one system dominates, depending on the environmental context.

The authors also suggest that the Cqs signal could be exploited therapeutically to dampen *V. cholerae* virulence. Quorum-sensing systems that trigger virulence have already been targets for therapies against several species, such as *Pseudomonas aeruginosa* — an opportunistic pathogen that causes infections in people with compromised immunity. Usually the goal is to block quorum sensing with a small-molecule inhibitor⁵. This is a difficult task, because such inhibitors must be specific, stable, easily deliverable to the infection site and able to out-compete the natural quorum-sensing signal for the target receptor.

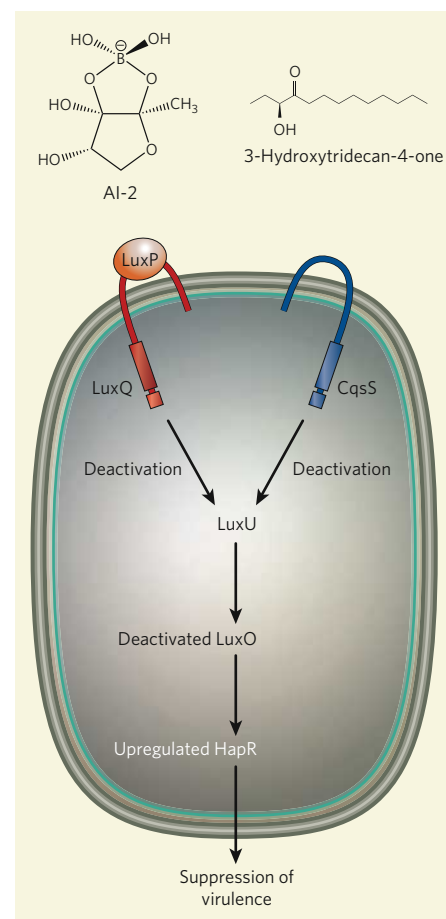


Figure 1 | Inhibiting virulence in *Vibrio cholerae*. In *V. cholerae* bacteria, the HapR regulator represses the expression of virulence genes. HapR expression is usually inhibited by the transducer protein LuxO, so that the bacteria are virulent. But *V. cholerae* emit two types of signal molecule that inhibit virulent behaviour in nearby *V. cholerae* bacteria. One of these, AI-2, is recognized by the LuxP receptor on the bacterial cell membrane. LuxP activates the LuxQ protein inside the cell, which deactivates the transducer protein LuxU. This prevents activation of LuxO, so that HapR activity is increased and virulence is suppressed. Higgins *et al.*⁴ show that the second signal molecule used by *V. cholerae* is 3-hydroxytridecan-4-one. This signal interacts with a putative receptor on the cell membrane (CqsS) that then deactivates LuxU, triggering the same signalling cascade described for AI-2 and LuxPQ.

But quorum sensing in *V. cholerae* may be unique. Unlike in many other species, in which quorum sensing induces virulence, in *V. cholerae* the response shuts virulence down, allowing the bacteria to escape the host and re-enter the environment. This means that the signals themselves could be used as therapeutics. The high specificity of 3-hydroxytridecan-4-one for *V. cholerae* certainly makes it an excellent candidate for drug development. In fact, in a proof-of-concept experiment, Higgins and colleagues⁴ show that their synthetic version of the signal does indeed terminate production of known virulence

factors in *V. cholerae*. But this idea raises possible public-health issues: the activation of quorum sensing in *V. cholerae* also induces active movement of the bacterium, potentially mobilizing the pathogen and encouraging the spread of infection from one person to another.

For several years, the repertoire of bacterial quorum-sensing signal molecules and receptors was thought to be rather limited and restricted to a few species. But recent studies have revealed an array of different signals, suggesting that we have only just scratched the surface of possible mechanisms. As new signals are identified and their use by bacteria is assessed, the list of quorum-sensing organisms will undoubtedly grow. We may

eventually reach a point at which bacteria that do not engage in quorum sensing are regarded as the exception, rather than the norm. The challenge now is not only to identify new systems, but also to make sense of why an organism would use one type of system over another. ■

Matthew R. Parsek is in the Department of Microbiology, University of Washington, Box 357242, Seattle, Washington 98195-7242, USA. e-mail: parsem@u.washington.edu

1. Miller, M. B., Skorupski, K., Lenz, D. H., Taylor, R. K. & Bassler, B. L. *Cell* **110**, 303–314 (2002).
2. Zhu, J. et al. *Proc. Natl Acad. Sci. USA* **99**, 3129–3134 (2002).
3. Camilli, A. & Bassler, B. L. *Science* **311**, 1113–1116 (2006).
4. Higgins, D. A. et al. *Nature* **450**, 883–886 (2007).
5. Sperandio, V. *Expert Rev. Anti-infect. Ther.* **5**, 271–276 (2007).

PALAEOCLIMATE

Slush find

Alan J. Kaufman

A coupled model of palaeoclimate and carbon cycling turns up the heat on the idea that Earth once became a giant snowball. It supports instead a milder 'slushball Earth' history — but piquant questions remain.

Sediments laid down in the oceans during the late Neoproterozoic era, between about 850 million and 542 million years ago, tell a dramatic story. They contain wildly varying abundances of the carbon isotope ^{12}C , which is typically incorporated into organic matter during photosynthesis. The pattern of excess ^{12}C in carbonates immediately above and below glacial deposits seems to indicate that photosynthesis on Earth came to a halt during a series of ice ages. These observations are a foundation of the 'snowball Earth' hypothesis^{1,2}: that, just before the first appearance of animals, Earth's surface might have been repeatedly frozen over, even at tropical latitudes.

Not necessarily so, say Peltier *et al.* on page 813 of this issue³. They apply basic ideas about the solubility of gases to a coupled model of climate and carbon cycling⁴ during the frigid late Neoproterozoic era. The results that emerge might explain the oscillatory carbon-isotope compositions of carbonates across the Neoproterozoic glacial cycles, without resorting to the hard-snowball model. Instead, they could lend support to a milder variation on the same theme — 'slushball Earth'.

The slushball and snowball models both predict ice sheets on continents near the Equator, but with markedly different extents of ice covering the oceans. In the snowball version, the frozen planet is completely blanketed, and reflects most of the Sun's warming rays back into space. Temperatures plummet and surface processes, including life, largely cease. Escape from the snowball state probably requires the build-up of volcanic carbon dioxide in the atmosphere over many millions of years,

resulting in torrential acid rain and the intense weathering of exposed rocks during the global thaw.

The slushball model⁵, by contrast, predicts open glacial oceans that would have constrained runaway refrigeration by allowing sunlight to warm the planet's surface, driving an active hydrological cycle⁶ and photosynthesis⁷ in exposed seas. The end of such an ice age need not have required extreme amounts of CO_2 in the atmosphere, nor have been delayed for millions of years.

Peltier and colleagues' new dynamic model³ shows how climate and atmospheric oxygen might have combined to prevent a runaway snowball Earth. As the oceans cool during ice ages, lower temperatures allow atmospheric gases such as oxygen to diffuse more readily into the deep sea, forcing the oxidation of abundant dissolved organic carbon, formed initially by photosynthesis in surface waters, to CO_2 . Released back to the atmosphere by this oceanic 'respiratory' process, the excess CO_2 would warm the planet and thereby end the glacial epoch.

What is particularly interesting about this model is that climate drives the carbon cycle (and so determines the stable levels of atmospheric CO_2). In the most recent ice ages, as well as for earlier interpretations of Neoproterozoic carbon-isotope anomalies⁸, the assumption has instead been the other way around. The crucial difference is that the Neoproterozoic carbon cycle was conceivably buffered by a marine pool of dissolved organic carbon that was orders of magnitude larger than that in the present-day oceans⁴.



A. J. KAUFMAN

Figure 1 | A soluble solution? The large (5–8-cm high) carbonate crystal fans (black to dark grey), which seem to grow out of the sea floor in this polished slab of a Neoproterozoic 'cap carbonate' from Brazil, suggest the presence of high concentrations of dissolved inorganic carbon in sea water after the ice ages, together with the rapid accumulation of sediments. These fans are draped by grey to white, fine-grained carbonates, which near the top become red, probably because they contain the iron-oxide mineral haematite (Fe_2O_3). The isotopic composition of such geological deposits is a focus of Peltier and colleagues' model interpretation³ of Neoproterozoic climate and carbon cycling.

A pertinent criticism of Peltier and colleagues' mathematical model is the uncertainty in its input parameters, in particular the assumption that levels of atmospheric oxygen were similar to those of today (around 21%). Biological⁹ and geochemical^{10–12} evidence indicates that oxygen levels were low throughout most of the Neoproterozoic, with a significant rise in breathable air around 550 million years ago — about the time animals first appeared on the planet. In that case, it seems likely that pervasive oxygenation of the atmosphere and the hydrosphere, including the vast pool of dissolved organic carbon, occurred millions of years after the extensive ice sheets of the Neoproterozoic had melted away. This rise, known as the Wonoka anomaly after the locality in South Australia in whose rocks it was first observed, is recorded in 550-million-year-old carbonates worldwide that are spectacularly rich in ^{12}C .

The coupled model also does not address certain hallmark geological features of the Neoproterozoic glacial episodes. These include the unexpected appearance of iron-bearing sediments in the glacial deposits, as well as the enigmatic 'cap carbonates' that lie immediately above them (Fig. 1). The co-occurrence of iron-oxide cements and glacial sediments implies that levels of soluble iron increased during

OCEANOGRAPHY

Siberian salt in the cellar

The Lomonosov ridge, a mountain range under the Arctic Ocean, gained unusual notoriety in August, when a Russian submarine expedition planted a rust-proof titanium flag there to reinforce the country's Arctic territorial claims.

Now Brian Haley and colleagues report in *Nature Geoscience* that the ridge furnishes evidence of Russia's past influence on the region — at least, on its ocean circulation (B. A. Haley *et al.* *Nature Geosci.* doi:10.1038/ngeo.2007.5; 2007). They study neodymium (Nd) isotopic ratios in marine sediments in a core of sediments drilled from the Lomonosov ridge near the North Pole, at a sea depth of 1,250 metres.

The core represents a historical sketch of Arctic oceanography over the past 65 million years. The authors' big news is that the ratio $^{143}\text{Nd}/^{144}\text{Nd}$ of deep Arctic water that is preserved in the sediments was consistently far higher in the Neogene period between 15 million and 2 million years ago than it is now,

indicating the influence of young, mantle-derived rock. In the past 2 million years, similarly high ratios are found only during short ice ages.

Haley *et al.* argue that the only credible source for such a signal is material from the Putorana basalts of the Central Siberian Plateau. But how did a surface-water signal from the Siberian coastal shelf sink to the great depths of the Lomonosov ridge?

The authors see the answer in the wider global climate, and the opening up of the Fram Strait 17.5 million years ago. This strait is the only deep passage for water to and from the Arctic, and runs roughly along the Greenwich meridian between Greenland and the Norwegian island of Spitsbergen. Its breaching allowed warm, saline waters to flow into the Arctic from the Atlantic. As these waters entered colder climes, evaporation increased, causing more precipitation and the growth of extensive floating ice shelves at northerly latitudes —



including along the Siberian coast.

As sea water freezes, it rejects salt. Haley *et al.* propose that the resulting denser, briny water that sank beneath the developing ice sheet carried Russian sediment to the Lomonosov ridge during the Neogene, as it similarly does off Antarctica today.

Waters from the North Atlantic drift have low neodymium ratios. The authors estimate that, to maintain the ratios of the Neogene core record after the initial influx following the Fram breach, the flow of warm water from the Atlantic into the Arctic could have been no more than half of that today. They suggest that the Atlantic conveyor belt must

at that time have stopped at a more southerly point than it does today. This is a well-established prediction for recent ice ages, which indeed also produce high neodymium ratios in the Arctic cores. But how it would have been maintained as an equilibrium state during the entire Neogene is unclear.

The model of an Arctic circulation dominated by seawater subduction off the Russian coast contrasts with today's picture, in which the Arctic is under a steady North Atlantic influence. In sketching the Arctic's sensitivity to past climate change, Haley *et al.* underscore its vulnerability to further change today.

Richard Webb

the ice age. As iron-bearing minerals such as haematite (Fe_2O_3) are remarkably insoluble in the presence of oxygen, large regions of the ocean must have been largely anoxic during the glacial periods — at odds with the authors' suggestion of progressive oxygenation. A whiff of oxygen would have caused an iron-rich sea to rust, potentially consuming much of the oxidant needed to convert dissolved organic carbon to CO_2 . Other potential sinks for oxygen, including weathering of the continents and the oxidation of volcanic gases, were similarly overlooked in the model exercise.

The cap carbonates are testament to the extreme build-up of carbonate anions (alkalinity) in sea water during the Neoproterozoic glacial episodes, and to their wholesale accumulation as carbonate minerals in the glacial aftermath. The oxidative respiration of organic matter produces CO_2 and also creates alkalinity, so Peltier and colleagues' open-ocean solution might also explain the ubiquitous presence of the cap carbonates. But as the authors acknowledge³, there are other possible oxidants that would work in an anoxic glacial ocean — sulphate, for example¹³. In the absence of free oxygen, sulphate-reducing bacteria could have occupied the water column, as they do in the Black Sea today, and could have fed on the standing pool of organic carbon, progressively raising the concentrations of inorganic carbon. At the same time, their metabolic activity would have released hydrogen sulphide

that, when combined with iron, would form the highly insoluble mineral pyrite (FeS_2). The resultant rain of pyrite to the sea floor might help to explain extreme sulphur-isotope variations that are notably present in the post-glacial cap carbonates¹⁴.

These texturally and isotopically distinct carbonates figure prominently in Neoproterozoic palaeoclimate interpretations. In Peltier and colleagues' model, the ^{12}C -rich cap carbonates reflect one stable state of the carbon cycle. But notably, isotopically similar carbonates also accumulated immediately before the ice ages^{7,15}. Depending on the timing of CO_2 release, the presence of these deposits can effectively neutralize the authors' proposed solubility hypothesis for the Neoproterozoic ice ages. Not only is more oxygen dissolved when the oceans get colder, so too is more CO_2 , which makes water acidic. Acidification of the oceans would have a profound effect on the preservation of carbonate deposited before or after the ice ages.

The variable accumulation of carbonate and iron-oxide-rich deposits across the glacial cycles is not necessarily incompatible with Peltier and colleagues' slushball model³. It could well reflect regional differences in seawater salinity and pH, as well as levels of soluble oxygen, sulphate, iron and dissolved organic and inorganic carbon in sea water. The idea of a self-limiting climate as expressed in their model is a tantalizing prospect, and an

important contribution to the debate. But our poor understanding of Neoproterozoic ocean dynamics and oxidation add great uncertainty to such mathematical models of Neoproterozoic climate and carbon cycling. ■

Alan J. Kaufman is in the Geology Department, University of Maryland, USA, and is currently on sabbatical at the Geologisch-Paläontologisches Institut, Westfälische Wilhelms-Universität Münster, Corrensstraße 24, 48149 Münster, Germany.

e-mail: kaufman@geol.umd.edu

1. Kirschvink, J. L. in *The Proterozoic Biosphere: A Multidisciplinary Study* (eds Schopf, J. W. & Klein, C.) 51–52 (Cambridge Univ. Press, 1992).
2. Hoffman, P. F., Kaufman, A. J., Halverson, G. P. & Schrag, D. P. *Science* **281**, 1342–1346 (1998).
3. Peltier, W. R., Liu, Y. & Crowley, J. W. *Nature* **450**, 813–818 (2007).
4. Rothman, D. H., Hayes, J. M. & Summons, R. E. *Proc. Natl Acad. Sci. USA* **100**, 8124–8129 (2003).
5. Hyde, W. T., Crowley, T. J., Baum, S. K. & Peltier, W. R. *Nature* **405**, 425–429 (2000).
6. Arnaud, E. & Eyles, C. H. *Sedim. Geol.* **183**, 99–124 (2007).
7. Olcott, A. N. *et al.* *Science* **310**, 471–474 (2005).
8. Kaufman, A. J., Knoll, A. H. & Narbonne, G. M. *Proc. Natl Acad. Sci. USA* **94**, 6600–6605 (1997).
9. Knoll, A. H. & Carroll, S. B. *Science* **284**, 2129–2137 (1999).
10. Fike, D. A. *et al.* *Nature* **444**, 744–747 (2006).
11. Canfield, D. E., Poulton, S. W. & Narbonne, G. M. *Science* **315**, 92–95 (2007).
12. Kaufman, A. J., Corsetti, F. A. & Varni, M. A. *Chem. Geol.* **237**, 47–63 (2007).
13. Hayes, J. M. & Waldbauer, J. R. *Phil. Trans. R. Soc. Lond. B* **361**, 931–950 (2006).
14. Hurtgen, M. T., Halverson, G. P., Arthur, M. A. & Hoffman, P. F. *Earth Planet. Sci. Lett.* **245**, 551–570 (2006).
15. Halverson, G. P. *et al.* *Geochim. Geophys. Geosyst.* **3**, 10.1029/2001GC000244 (2002).

OBITUARY

Arthur Kornberg (1918–2007)

Passionate biochemist with a love for enzymes.

Arthur Kornberg was one of the greatest biochemists of the twentieth century. His career spanned more than 60 years, and such has been the impact of his work on modern biomedical science that his influence will endure for decades.

Kornberg was born in Brooklyn, New York, to a family of modest means. He had no exposure to science as a child; the young Arthur collected matchbook covers, the dominant flora of the Brooklyn streets, rather than the insects usually associated with a scientifically inclined youngster. He nonetheless opted to study chemistry and biology, and received his bachelor's degree from the City College of New York, one of three students from the class of 1937 destined to receive the Nobel prize (Herbert Hauptman and Jerome Karle were the others).

He went on to earn an MD from the University of Rochester, New York, where he wrote his first research paper on a mild form of jaundice he discovered in himself and classmates. For a short period he was a (self-described) attentive physician, first in New York, and then in the US Navy. But it was a transfer to the Public Health Service in 1942, which landed him at the National Institutes of Health (NIH), that was the turning point of his career.

At the NIH, and during research with Severo Ochoa at the New York University School of Medicine, and with Carl and Gerty Cori at Washington University School of Medicine, Kornberg turned to the study of enzymes, which he loved and considered the vital force in biology. He initially analysed enzymes involved in the synthesis of nucleotides and related small molecules. Nucleotides are the building-blocks of the molecule of heredity, DNA. Kornberg soon turned to the search for an enzyme that could synthesize this polynucleotide chain. His inspiration was not the 1953 Watson–Crick structure of DNA, but rather the Coris' work on the enzymatic synthesis of other biological macromolecules.

Kornberg had taken on a daunting task. A guiding principle of his research, however, was that a persistent biochemist could reconstitute any cellular event in a test tube. Moreover, he felt that the biochemist had the advantage of being able to optimize the reaction without the burden of working under the constraints imposed by cells. Kornberg set up a reaction containing DNA, radioactive thymidine (a nucleoside) and protein fractions obtained by breaking open

Escherichia coli, an intestinal bacterium. Initially, only a tiny fraction of the labelled thymidine was converted into a form that seemed to be a DNA-like molecule. But, with this glimmer of activity (and many controls demonstrating reaction specificity) Kornberg persevered, and eventually found the proper form of the building-blocks needed for efficient synthesis.

Late in 1957, Kornberg's manuscripts describing the test-tube synthesis of DNA from precursor molecules by an enzyme — DNA polymerase — were rejected by *The Journal of Biological Chemistry*. Reviewers thought it premature to call the product 'DNA'. In the spring of 1958, however, a new editor stepped in and accepted the papers. Less than two years later, Kornberg's discovery, and the independent work on the synthesis of RNA polymers by Ochoa, were recognized by the Nobel Prize in Physiology or Medicine.

The enzymology of DNA replication remained Kornberg's passion for the next 30 years. His approach was to build up increasingly complex biochemical reactions that recapitulated cellular DNA replication processes. In 1967, he successfully replicated the DNA of a small virus, a feat that received much attention as it was (incorrectly) reported that he had created "life in a test tube". This was nonetheless a seminal achievement that defined many of the principles governing DNA replication of all genomes.

For example, the concept that special enzymes are needed to start DNA chains, which DNA polymerase then elongates rapidly and with exquisite fidelity, arose from this work. Likewise, Kornberg's work provided insight into the function of origins of replication, the sequences on a chromosome that direct the start of a cycle of DNA replication. In the 1980s, after ten years of effort, Kornberg's group reconstituted the replication of a DNA molecule initiated using a true origin 'cloned' from a cellular chromosome. The resulting model of the steps required for initiation of DNA replication at a chromosomal origin remains the paradigm for investigating replication in more complex organisms today.

From 1990, Kornberg's lab focused on the enzymes involved in polyphosphate metabolism. He continued this work until a week before his death on 26 October.

Kornberg's impact on modern biology has been enormous. He spearheaded the purification and characterization of many



BETTMANN/CORBIS

of the enzymes that have become essential tools of genetic engineering. He was also a visionary leader, who had an immense influence on the education of scientists. In 1959, he founded the biochemistry department at Stanford University. There, he forged a remarkable community, in which faculty limited the size of their research groups, pooled grants, and shared space and equipment.

He also successfully integrated his work and family life, bringing his children Roger, Tom and Ken, and then his grandchildren, to the lab and to conferences. His first wife, Sylvvy (pictured here with him), was a gifted biochemist and worked alongside him for years. To Arthur's great pride and joy, Roger became a colleague at Stanford and, in 2006, received the Nobel Prize in Chemistry.

Then there is Arthur's extended scientific family. As a mentor he was passionate about both his science and his people, and held us to the same exacting standards as he held himself. Although fundamentally impatient, he was never too busy or too famous to discuss our data, edit our manuscripts (again, and again, and again), or to critique our presentations. He drove us crazy with his (not so gentle) homilies. "Keep your lab notebook with more accuracy than your cheque-book." "Only four lines per graph." "Don't go home until you find your enzyme activity." "Time is your most important resource." But we learned, treasure the memories, and now torture our own students with 'Kornbergisms'. I think Arthur would be pleased.

Tania A. Baker

Tania A. Baker is in the Department of Biology, Massachusetts Institute of Technology, and the Howard Hughes Medical Institute, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.
e-mail: tabaker@mit.edu

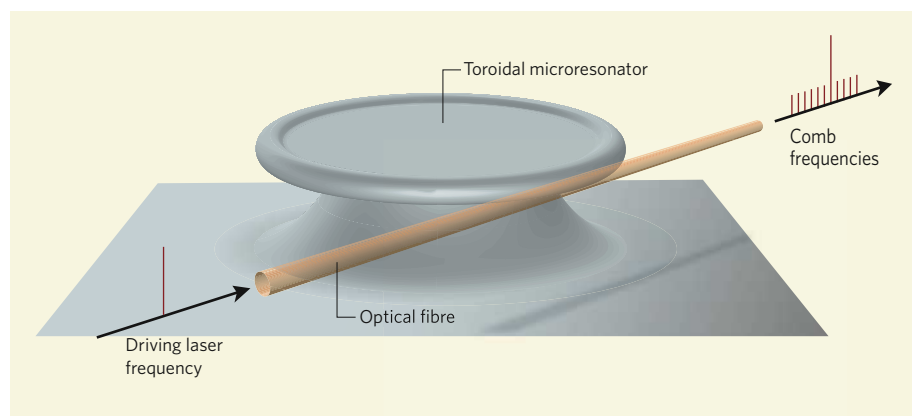


Figure 1 | Unfurling the standard. The high 'quality factor' of the toroidal microresonator used by Del'Haye *et al.*⁸ allows a driving laser field to be held for a long time, greatly increasing the light intensity to the point at which the response of the resonator's glass becomes nonlinear. When a continuous wave laser, corresponding to a single, well-known frequency spike (left), is launched into this microresonator, its light is converted into a regularly spaced comb of frequencies (right) through its interaction with the wave modes of the resonator cavity. Such combs allow frequencies to be measured accurately over a wide range of frequencies (in this case optical), as each 'tooth' acts as a reliable frequency standard.

The philosophy changed in the year 2000, with the advent of femtosecond (10^{-15} s) optical frequency combs^{1,2}. These combs are produced by mode-locked lasers whose light comes in short, sharp bursts of femtosecond duration. A train of such short pulses can be decomposed into light at the laser's offset frequency (the rate at which its phase evolves) plus integer multiples of the laser's repetition rate (the number of pulses produced per second). The shorter the pulse, the wider the frequency spectrum spanned by these frequency lines. Under the right conditions, a spectrum spanning hundreds of terahertz — broad enough to cover the entire visible spectrum, and thus look white to the eye — can be produced.

The breadth of the band is crucial, because most methods for measuring the offset frequency use a self-referencing technique that requires a spectrum spanning an octave (a factor of two in frequency)¹⁰. The great advantage of the femtosecond-laser approach is that it locks the phases of all comb lines together. By contrast, the electro-optic modulator locks the phases only of adjacent comb lines, allowing phase variations to build up towards the edges of the comb, which limits the precision of any frequency measurement there.

In some ways, the new approach to optical-frequency generation taken by Del'Haye *et al.*⁸ is similar to older radio-frequency techniques. Like the radio-frequency approach, it takes a sinusoidal input signal — the output of a continuous-wave laser at a wavelength of 1,550 nanometres — and couples it to a nonlinear response medium, in the form of a toroidal microresonator¹¹. This microresonator stores up the laser light, greatly increasing its intensity as more and more light enters such that nonlinear wave mixing occurs. The frequencies of the output lines from the microresonator are equally spaced on either side of the frequency of the input laser light, with a spacing determined by the properties of the microresonator (Fig. 1).

The nonlinear generation of a pair of frequency lines either side of an optical signal is not new. The generation of tens of them is, and Del'Haye *et al.* achieve it by using a microresonator with a very large 'quality factor' — a measure of the sharpness of its resonant response. This not only enhances the resonator's nonlinearity but also produces a cascading effect, with each frequency line generating the next, just as in the electro-optic comb.

For a spectral comb to be useful as a frequency reference, the spacing of the comb lines must be perfectly even, and the lines themselves must be narrow. Conservation of energy and momentum show why the spacing of lines in a toroidal microresonator must be even: in the initial nonlinear response, two 'pump' photons of frequency ν_p are destroyed to generate one photon in each of the sidebands. These have frequencies ν_+ and ν_- , which are related by $2\nu_p = \nu_+ + \nu_-$, according to energy conservation. This allows a continuum of frequencies; momentum conservation, which expresses itself as the requirement that the phase evolution of the two lines be matched, then selects out one pair of frequencies for the first two lines. Once the initial pair of lines is generated, the cascade that then builds up preserves their spacing.

By comparing the comb generated by a microresonator to a femtosecond comb, Del'Haye and colleagues demonstrate that the spacing is indeed regular, to better than one part in 10^{-17} , rivalling the quality of femtosecond combs¹². The implication, although not yet definitely proved, is that the comb lines are narrow, and thus phase noise is not building up significantly in the nonlinear cascade.

An obvious advantage of the new device is its small size and potential for integration with optical-fibre technology. But the offset frequency of the comb being determined by the frequency of the pump laser is a double-edged sword: if the laser frequency is known

and stable, it is an advantage; if it is unstable, it is a disadvantage. If the frequency of the pump laser is not known, it can be determined using the same self-referencing technique used to determine the offset frequency in a femtosecond comb. This technique is easiest to apply if the width of the comb is pushed to cover an octave (in Del'Haye and colleagues' set-up⁸ it is already close).

Probably the biggest obstacle to using the authors' comb for optical-frequency metrology is the spacing of its teeth, which is around 1 terahertz. The ideal comb spacing is of the order of the detection bandwidth, which is typically a few gigahertz — a factor of around 1,000 smaller. Femtosecond combs produced by mode-locked lasers typically have a tooth spacing of 100 megahertz to 1 gigahertz. This very tight spacing is also less than ideal, because it reduces the power per comb line, but the spacing can be increased by filtering out a subset of the comb lines. Lowering the spacing of the microresonator combs to this value will be a challenge: it would require increasing the diameter of the toroid to about 0.5 mm, while preserving its extraordinarily high quality factor and simultaneously increasing the laser power to around 5 watts to maintain the nonlinear effect. Overcoming such problems will make for interesting times for this exciting new comb technology.

Steven T. Cundiff is at JILA, National Institute of Standards and Technology, University of Colorado, Boulder, Colorado 80309-0440, USA.
e-mail: cundiffs@jila.colorado.edu

1. Jones, D. J. *et al.* *Science* **288**, 635–639 (2000).
2. Holzwarth, R. *et al.* *Phys. Rev. Lett.* **85**, 2264–2267 (2000).
3. Cundiff, S. T. & Ye, J. *Rev. Mod. Phys.* **75**, 325–342 (2003).
4. Udem, Th., Holzwarth, R. & Hänsch, T. W. *Nature* **416**, 233–237 (2002).
5. Ma, L. S. *Opt. Photon. News* **18**(9), 42–47 (2007).
6. Thorpe, M. J., Moll, K. D., Jones, R. J., Safdi, B. & Ye, J. *Science* **311**, 1595–1599 (2006).
7. http://nobelprize.org/nobel_prizes/physics/laureates/2005/index.html (2005).
8. Del'Haye, P. *et al.* *Nature* **450**, 1214–1217 (2007).
9. Kourogi, M., Nakagawa, K. & Ohtsu, M. *IEEE J. Quantum Electron.* **29**, 2693–2701 (1993).
10. Telle, H. R. *et al.* *App. Phys. B* **69**, 327–332 (1999).
11. Armani, D. K., Kippenberg, T. J., Spillane, S. M. & Vahala, K. J. *Nature* **421**, 925–928 (2003).
12. Ma, L.-S. *et al.* *Science* **303**, 1843–1845 (2004).

Correction

The photograph accompanying the obituary of Arthur Kornberg by Tania A. Baker (*Nature* **450**, 809; 2007) was inverted left-to-right. Here is the picture in the correct orientation. It shows Dr Kornberg, his first wife Sylvy, and a model of the DNA double-helix (now right-handed, not left-handed).



Q&A



JUXTAPOSE/GETTY

SUSTAINABLE TECHNOLOGY

Green chemistry

Martyn Poliakoff and Pete Licence

Modern life depends on the petrochemical industry — most drugs, paints and plastics derive from oil. But current processes for making chemical products are not sustainable in terms of resources and environmental impact. Green chemistry aims to tackle this problem, and real progress is being made.

Why is chemical manufacture becoming unsustainable?

There are two main reasons. The first is that most chemical products — from perfumes to plastics to pharmaceuticals — are based on carbon, which currently is supplied by Earth's finite petroleum feedstocks. Alternative carbon sources do exist — for example, coal was the basic feedstock for chemical production before oil, and could be used again. But readily accessible coal is also in limited supply, and the conversion of coal into fine chemicals requires catalysts based on metals that are themselves becoming scarce. The second, equally pressing issue is waste. In general, industrial chemical processes generate large amounts of waste, the safe disposal of which imposes an increasing burden on the environment.

Can't these problems be dealt with using existing technology?

Unfortunately not. The chemical industry has made great improvements in the efficiency of its processes, but finite natural resources will inevitably limit our ability to manufacture chemical products for the increasing global population. Unless there is radical innovation, the industry will eventually struggle even to supply existing consumers. The challenge of green chemistry for both academics and industrialists is to devise sustainable strategies that

meet the demand for chemical products from an ever-increasing population.

What aspects of manufacturing does green chemistry address?

All of them. Working through the arc of an industrial process, the first challenge is to identify renewable feedstocks. The current front-runners are non-food plants, in which case chemists must find effective ways of converting the whole plant into useful products. Next, the reactions involved in making chemical products must be devised to minimize environmental impact. For example, many traditional catalysts are based on metals, which can be toxic or scarce; non-metallic catalysts must therefore be developed, perhaps based on organic compounds (organocatalysts) or on enzymes that have been modified to perform useful reactions. Engineering is also crucial — industrial processes and reactors must be designed to maximize efficiency and reduce waste. Improved analytical techniques are necessary to monitor the fate of potentially harmful chemicals in reactions and in the environment. And finally, the impact of chemicals on the environment can be reduced by finding replacements with reduced toxicity and increased biodegradability compared with existing mass-produced compounds.

How can you tell if a chemical process is green?

This question puzzled chemists for some time. A standard set of 12 principles has emerged that can be used to assess any process (Fig. 1). These address several issues, including the amount of waste produced; the number of sequential chemical steps (the fewer the better); the use of catalysts (which is to be encouraged, because catalysts are needed only in small amounts and reduce the quantities of reagents used); and the toxicity of the products. It isn't expected that new chemical processes should always satisfy all 12 principles, but the checklist does provide a rough idea of whether one process is greener than another.

Is it possible to quantify how green a process is?

Several methods have been proposed, but the easiest thing to measure is the amount of waste generated. From the balanced chemical equation for a process, the theoretical quantity of waste can be calculated per unit mass of starting material, assuming that the yield of product is 100%. But in reality, most reactions give lower yields, and the actual amount of waste is higher than the theoretical value. To address this issue, processes are assessed using their 'E-factor' — the ratio of the mass of waste to that of the product. All processes should aim

for the lowest possible E-factor; for truly green processes, the E-factor should be zero.

So which manufacturing processes are currently the worst, according to their E-factors?

E-factors often throw up surprises. Large-scale manufacturing processes for bulk chemicals are generally perceived as being bad for the environment, compared with the relatively small-scale operations of the pharmaceutical sector. But E-factors for bulk chemical manufacturing are typically much less than 5 — even though the volumes produced are so high that the amount of waste is very large. By contrast, E-factors in the pharmaceutical industry can be greater than 100; for example, when the antidepressant sertraline was launched, 250,000 litres of solvent were needed for each 1,000 kilograms of product. Applying the principles of green chemistry reduced solvent usage tenfold.

Can't waste be completely eliminated from chemical processes?

Some waste is unavoidable, because energy is required to break the bonds in the starting materials of a reaction. If the energy input is not balanced by the energy generated from bonds forming in the product (which is often the case), then extra bonds must be created, usually in a by-product. Some by-products can be used as feedstocks for further reactions — for example, nitrous oxide (N_2O) generated in the production of nylon can be used as an oxidant to convert benzene into phenol, a bulk chemical with many applications. But many by-products are insufficiently reactive or too diluted to be recycled economically. In that case, the most sensible treatment for organic waste is careful incineration, so that at least its energy content can be exploited. But often, by-products do not form the bulk of the waste — solvents do.

So, are solvents high on the green-chemistry agenda?

They are a top priority. Many solvents are flammable or toxic, and most are volatile organic compounds that contribute to atmospheric pollution. Solvents are necessary for most reactions because they aid mixing, transfer heat and sometimes control the reactivity of reagents. But the largest quantities of solvents are used in the isolation, separation and purification of materials. If reactions could be optimized to prevent the formation of by-products, then purification processes would be eliminated (or at least minimized), greatly reducing the amounts of solvents required for the overall process. Finding environmentally benign solvents would be even better.

What are the alternatives to traditional solvents?

Some reactions don't need any solvent, although the lack of a medium for heat transfer can make it difficult to stop certain reactions from potentially disastrous over-heating.

But the most useful alternatives to traditional solvents are supercritical CO_2 (where the gas is compressed until it is nearly as dense as a liquid), ionic liquids (organic salts that are liquid at room temperature) and water.

Why not just replace all organic solvents with water?

One problem is that most organic compounds are insoluble in water. Despite this, a surprising number of organic reactions can be carried out in water; for example, propene (a hydrocarbon gas) can be reacted catalytically with carbon monoxide and hydrogen to make butanal — a reactive compound used as a starting material for a wide range of products. But often the main difficulty with aqueous systems is recovering products from them; evaporating large amounts of water is energy-intensive. One also has to keep the big picture in mind — claims that reactions are green because they are performed in tiny volumes of water are misleading if the purification process requires much greater quantities of organic solvents. It's like arguing that air travel is environmentally friendly because the passengers walk from the terminal to the plane.

What is so good about supercritical CO_2 ? Wouldn't its widespread use contribute to climate change?

The solubility of materials in supercritical CO_2 can be varied merely by changing the pressure of the gas. This opens up separation and purification methods that are not possible with conventional solvents. For example, a modest pressure of CO_2 can cause some compounds dissolved in water to separate out without the need for distillation, which could greatly simplify industrial-scale processes. Supercritical CO_2 is also revolutionizing the chromatographic purification of many active pharmaceutical ingredients, because the high rates of diffusion of compounds in CO_2 make the separation more effective than with conventional solvents. Furthermore, supercritical CO_2 is non-toxic — it has been used for many years to decaffeinate coffee. It wouldn't contribute to climate change, because the CO_2 used is a by-product from other processes. The downside is that compressing CO_2 to a supercritical form is energy-intensive.

Why are ionic liquids hailed as green solvents?

The great advantage of ionic liquids is that they undergo very little evaporation and so are not lost to the atmosphere. The biggest issue is their potential toxicity. Only a few have been tested so far and, as might be expected, some are as toxic as conventional solvents whereas others are relatively innocuous. But as with water, one must be cautious about labelling a reaction as 'green' just because it uses an ionic liquid as the solvent. Combinations of ionic liquids with other green solvents could be especially useful — for example, supercritical

- P**revent wastes
- R**enewable materials
- O**mit derivatization steps
- D**egradable chemical products
- U**se safe synthetic methods
- C**atalytic reagents
- T**emperature, pressure ambient
- I**n-process monitoring
- V**ery few auxiliary substances
- E**-factor, maximize feed in product
- L**ow toxicity of chemical products
- Y**es, it is safe

Figure 1 | The principles of green chemistry. Green chemical processes adhere to 12 principles, shown here in a simplified version to form a mnemonic. Catalytic reagents reduce the amount of chemicals needed in a reaction; in-process monitoring allows harmful substances to be detected and eliminated; auxiliary substances are those that don't take part in the chemical reaction, such as solvents or separating agents; and the E-factor is the mass of waste generated in a process divided by the mass of product. (Figure taken from S. L. Y. Tang, R. L. Smith and M. Poliakoff *Green Chem.* 7, 761–762; 2005.)

CO_2 has been used to pass reactants through an ionic liquid that contains a dissolved catalyst and to extract the products in a continuous process. Catalysts are often very expensive, and trapping them in the ionic liquid prevents them from being lost.

Can any reactions be performed in green solvents that weren't possible in traditional solvents?

Green solvents have been a little disappointing in this respect, as few examples of such reactions have been described. But there are many cases in which known reactions can be carried out more efficiently, including some industrially useful ones. For example, reactions of organic compounds with hydrogen work well in supercritical CO_2 , and certain alkylation reactions — in which hydrocarbon groups are attached to aromatic compounds — excel in ionic liquids. Both of these types of reaction are widely used in the manufacture of chemical products.

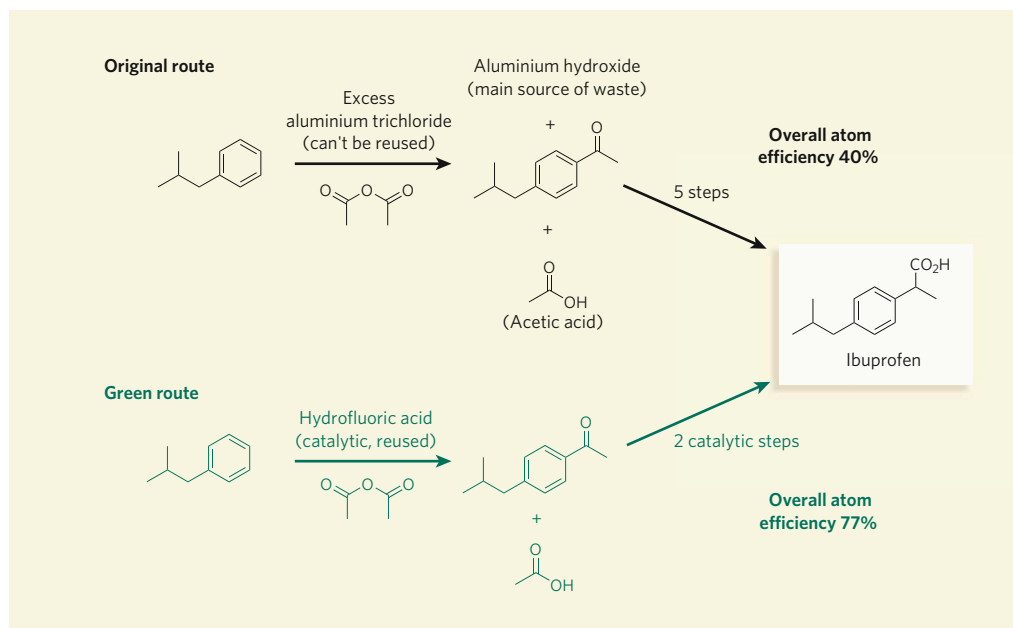
Doesn't chemical engineering have a role to play in green chemical processes?

Engineering chemical processes is just as vital as developing reactions. An exciting aspect of green chemistry is that it brings chemists and chemical engineers together. Green solvents open up new design options for reactors, and the need for quantitative data on reactions in these media is leading to renewed interest in physical organic chemistry. Chemistry and chemical engineering departments of universities are collaborating closely to address these needs.

Box 1 | Green ibuprofen

Ibuprofen is the main active ingredient in many over-the-counter painkillers. It was first prepared and patented in 1961. The original synthetic route involved six consecutive steps and had an overall atom efficiency of just 40% — of the mass of all the atoms going into the process, 60% ended up in waste products. This route was used to manufacture the drug until the patent expired in 1984. If ibuprofen were still made in this way, more than 20,000 tonnes of waste would be generated annually.

In the early 1990s, the BHC company (now part of BASF) redesigned the synthetic route using many of the principles of green chemistry. The power of its approach is demonstrated by comparing the first steps of the two routes (shown here). The same chemical transformation occurs, but it is achieved in very different ways. Originally, aluminium trichloride was required to promote the reaction. This generated aluminium hydroxide, which was filtered off as a cake of solid waste. To get a high-yielding reaction, the aluminium trichloride was needed in excess (in larger quantities than the theoretical amount suggested by the reaction equation), adding to the waste problem.



But in the green route, hydrofluoric acid is used to promote the reaction instead of aluminium trichloride. Because it is used as a catalyst, only a small amount of the acid is required; even better, it is recycled and used for making the next batch of product. In this way, a large amount of solid waste was eliminated from the process.

The green route then adopts a synthetic strategy that is quite different from the original process,

so that only two more steps are required (compared with five more needed in the original process). Each step eliminated reduces the resources used and the waste generated. Both the remaining steps in the green route are catalytic — unlike those in the original synthesis — further reducing waste.

In this way, the overall atom efficiency of the green process was increased to an impressive

77%. In principle, the efficiency could be improved to more than 99% by recycling the by-product of the first reaction (acetic acid, which can be easily converted to acetic anhydride, a reagent for the first step). The green route thus produces more ibuprofen in less time and using less energy than the original process — which means cheaper products for the consumer with increased profits for the manufacturers. **M.P. & P.L.**

How has re-engineering processes changed things?

One obvious improvement is the switch from batch reactions to continuous processing — which is a bit like replacing baths with showers. For a given mass of chemicals, a continuous reactor can be much smaller than the corresponding batch reactor because smaller amounts of material are undergoing reaction at any given time. And higher reaction rates can be achieved without overheating, as product and reacted materials are being continuously removed. Continuous processing is not only safer, but can often give a higher-purity product. It also allows manufacturing to be more flexible — small batches of compounds can be prepared on demand, rather than having to be made in large amounts at a time, which then have to be stored until they're needed.

Can the chemical industry break free from petroleum as a basic feedstock?

The only renewable source of carbon is biomass derived from plants. But biomass has a different chemical composition from oil (it contains more oxygen), so new reactions will have to be developed to turn it efficiently into useful compounds. This issue will have to be

addressed by the chemical industry, although biological approaches — such as using enzymes to convert biomass into chemicals — also show promise. Other obvious problems with biomass include the seasonal nature of crops, possible conflicts with food supplies and the variability of its composition — different processes may need to be devised to convert different kinds of biomass into chemicals.

How does industry view green chemistry?

Initially, there was little appreciation of the potential benefits. But industrialists have become more interested as it has become clear that green processes can be more profitable than traditional ones (Box 1). For example, enzymatic reactions have completely displaced conventional catalysis as a low-cost option in the manufacture of several generic pharmaceuticals. The globalization of the chemical industry means that it requires only a few committed individuals to trigger substantial changes. A relatively small group of industrial-process chemists has achieved a considerable reduction in the volume of toxic solvents used in pharmaceutical manufacture across the world, by producing a simple toolkit that identifies less harmful replacements.

What is the real intellectual challenge for the future?

Perhaps the greatest challenge facing green chemists is the eventual elimination of all environmentally harmful chemical products. In other words, when designing compounds for a particular application, how can we ensure from their conception that they have low toxicity and rapid biodegradability while retaining their desired effect? Chemists are still a long way from being able to predict the properties — both chemical and biological — of compounds on the back of an envelope. Reaching that point is a daunting task, but it will inspire the next generation of chemists. ■

Martyn Poliakoff and Pete Licence are in the Schools of Chemistry and Chemical Engineering, University of Nottingham, Nottingham NG7 2RD, UK. e-mails: martyn.poliakoff@nottingham.ac.uk; peter.licence@nottingham.ac.uk

FURTHER READING

www.epa.gov/greenchemistry
www.rsc.org/Publishing/Journals/gc/index.asp
Sheldon, R. A. *Green Chem.* **9**, 1273–1283 (2007).
Constable, D. J. C. *et al.* *Green Chem.* **9**, 411–420 (2007).
Beckman, E. J. *J. Supercrit. Fluids* **28**, 121–191 (2004).
Wasserscheid, P. & Welton, T. (eds) *Ionic Liquids in Synthesis* 2nd edn (Wiley-VCH, Weinheim, 2007).

Effective global conservation strategies

Arising from: R. Grenyer *et al.* *Nature* **444**, 93–96 (2006).

Using data on the global distribution of mammal¹, bird² and amphibian³ species, Grenyer *et al.*⁴ conclude that planning based on individual taxa does not provide efficient solutions for the conservation of other taxa. They also report that the performance of existing global conservation strategies—endemic bird areas⁵, biodiversity hotspots⁶ and global 200 ecoregions⁷—in representing those taxa is often no better (and in some cases worse) than random. I argue here that the methodology used by Grenyer *et al.* was not appropriate for purported globally comprehensive analyses. Focusing on analyses of rare species as an example, I demonstrate how the data actually reveal substantial cross-taxon surrogacy and good performance of existing global conservation strategies.

Grenyer *et al.* describe the scope of their findings as ‘global’, but their analyses were only performed on subsets of their data. For example, when investigating the congruence between rare birds and rare mammals, they did not obtain a global correlation (across ~19,500 land-grid cells) but a correlation across cells holding rare species of birds or mammals (~5,800 cells). Hence, they excluded ~13,700 cells in which there was perfect agreement between rare birds and rare mammals, in that both were absent. Consequently, the correlation coefficient reported ($r = 0.48$) underestimates the true global correlation ($r = 0.58$). The same applies to all other r values reported, including those in their Fig. 2a–c (ref. 4).

The analyses of species numbers expected by chance in areas of variable size (random curves in their Fig. 4; ref. 4) suffer from the same methodological bias: for instance, the curve in their Fig. 4b is not for a true random selection, but for a random selection across just the ~6,000 cells containing rare species. The true global random curve is much shallower (Fig. 1). Globally, endemic bird areas, hotspots and global 200 ecoregions perform substantially better in

representing rare vertebrate species than would be expected by chance (Fig. 1). Given that hotspots were selected on the basis of plant endemism, this provides evidence for surrogacy of restricted-range plants in representing rare vertebrates.

Grenyer *et al.* investigated cross-taxon surrogacy by counting how many target species are represented in minimum complementary sets selected for a particular surrogate taxon (their Table 1; ref. 4). For rare species, they found values ranging from 22.5% to 77.9% and concluded that surrogacy is low. However, these values alone are not informative: they need to be compared with what would be expected by chance, and what the maximum possible representation is, in an area of the same size⁸. Figure 1 provides this information for when the target is the representation of rare species across the three groups. It shows that minimum sets representing rare mammals, rare birds or rare amphibians individually represent substantially more overall rare species than would be expected by chance. Furthermore, these minimum sets are noticeably close to the maximum representation possible, which is indicative of a high degree of surrogacy.

Analyses of rare species are the most disrupted by the methodological problems described here, but the other two groups analysed by Grenyer *et al.* (all species and threatened species) are also affected.

In conclusion, the analyses in Grenyer *et al.* suffer from a systematic methodological bias that does not allow the results to be compared with the maximal possible representation. The prospects for global conservation planning are, in fact, positive, not dismal as portrayed⁴. It is true that better results will be obtained when high-resolution data become available for all taxa we aim to conserve. Nonetheless, at least for the terrestrial realm, good progress can be, and has already been, achieved by conservation planning based on existing data.

METHODS

The following databases were used: ADHoC database of geographic ranges of birds², owned and developed by the NERC Avian Diversity Hotspots Consortium; global mammal database^{1,4}, owned and developed by J. Gittleman; Global Amphibian Assessment^{3,9}, developed by SSC-IUCN, CABS-CI and NatureServe. Rare species are those in the lower quartile of the range distribution of each taxonomic group⁴. Optimizations were achieved with the GNU Linear Programming Kit package.

Ana S. L. Rodrigues¹

¹Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

e-mail: aslr2@cam.ac.uk

Received 24 July; accepted 25 September 2007.

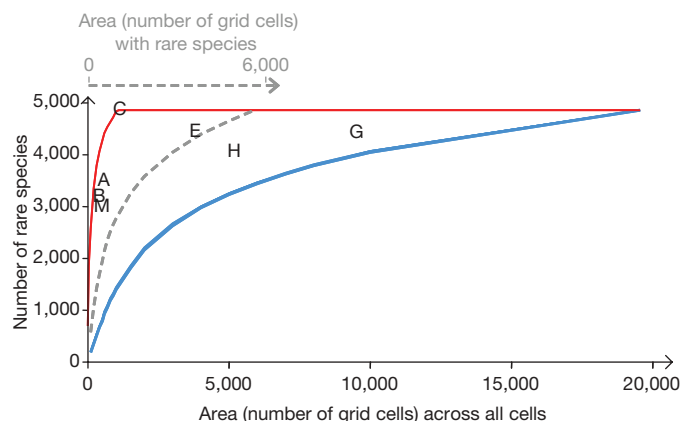


Figure 1 | Relative performance of different types of priority network in capturing rare species (see Fig. 4b of Grenyer *et al.*⁴). Minimum complementary set representing each species at least once of rare mammals (M), rare birds (B), rare amphibians (A), and rare mammals, rare birds and rare amphibians combined (C). Position is also shown for biodiversity hotspots (H), endemic bird areas (E) and global 200 ecoregions (G). Performance is evaluated by comparing the position of these points for equivalent areas with the maximum number of rare species that can be represented (red line) and the number of rare species expected by chance (blue: 95% confidence range for randomly selected sets of cells; 100 replicates). The dashed line is the random line from Grenyer *et al.*⁴, which corresponds to a random selection across just the sites holding rare species, plotted on the dashed grey x axis, top. In Fig. 4 of Grenyer *et al.*⁴, the position of the random lines therefore cannot be compared with the positions of points E, H and G as they are plotted on different axes (the positions of points M, B, A and C are the same on both axes).

1. Sechrest, W. *Global Diversity, Endemism and Conservation of Mammals*. Thesis, Univ. Virginia (2003).
2. Orme, C. D. L. *et al.* Global hotspots of species richness are not congruent with endemism or threat. *Nature* **436**, 1016–1019 (2005).
3. Stuart, S. N. *et al.* Status and trends of amphibian declines and extinctions worldwide. *Science* **306**, 1783–1786 (2004).
4. Grenyer, R. *et al.* Global distribution and conservation of rare and threatened vertebrates. *Nature* **444**, 93–96 (2006).
5. Stattersfield, A. J., Crosby, M. J., Long, A. J. & Wege, D. C. *Endemic Bird Areas of the World: Priorities for Biodiversity Conservation* (BirdLife International, Cambridge, UK, 1998).
6. Mittermeier, R. A., *et al.* *Hotspots Revisited: Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions* (CEMEX, Conservation International and Agrupación Sierra Madre, Mexico City, Mexico, 2004).
7. Olson, D. & Dinerstein, E. The Global 200: Priority ecoregions for global conservation. *Ann. Mo. Bot. Gard.* **89**, 199–224 (2002).
8. Ferrier, S. Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? *Syst. Biol.* **51**, 331–363 (2002).
9. IUCN, Conservation International and NatureServe. *Global Amphibian Assessment* (<http://www.globalamphibians.org>) (2006).

doi:10.1038/nature06374

Grenyer et al. reply

Replying to: A. S. L. Rodrigues *Nature* 450, doi:10.1038/nature06374 (2007).

Rodrigues¹ criticizes our demonstration² of low congruence in the global distributions of rare and threatened vertebrates on the grounds that we excluded locations where species counts were zero from our analyses. In practice, this makes no substantive difference to our conclusions. Some sample locations are not inhabited by any organisms of interest: such locations can inflate measures of covariation and association because their values for parameters of interest (in this case, zero counts of species) are identical. This bias has long been known (as 'the double-zero problem'³); many ecological techniques and studies exclude double-zero data for this reason.

We are happy to clarify that we excluded double-zero data. We concur with Rodrigues¹ that including double-zero data increases the observed correlation coefficient (r). Indeed, it cannot fail to do so: r becomes a function of the ratio of double-zero to non-zero data. For example, given our 19,564 terrestrial grid cells, if fewer than 13,043 cells (two-thirds of the world's surface) are filled with perfectly negatively correlated richnesses and the rest with double zeroes, a positive correlation is observed. In any such case, including double-zero data would be to predetermine the result: our comparison of rare mammals and amphibians involved only 3,184 non-zero cells. This is sufficient justification for our choice of method.

Treating all cells as informative, as Rodrigues proposes, becomes analogous to the analysis of binary presence-absence data: it examines associations between cells that have any species of interest, rather than the relative diversities of those cells. Such questions are better addressed by explicitly converting species counts to binary presence-absence data before analysis. The analyses would rely more on the

veracity of data on absence, and the metrics of association will depend on the relative abundance of zero to non-zero data.

Rodrigues also criticizes our description of network performance; our two contrasting methods are shown in Fig. 1. Our definition equates performance with the trade-off between network size and species capture. Rodrigues prefers a measure that quantifies the effectiveness of a network given its size. Both measures describe different aspects of reserve network performance. We prefer our definition in this case because we regard the comparison of absolute network sizes to be important. For example, under Rodrigues' definition, the optimal network for rare mammals and the global 200 scheme (points M and G in Rodrigues' Fig. 1; ref. 1) have an effectiveness of around 70% and 50%, respectively, despite differing in area by more than an order of magnitude.

We agree that good progress has been made in conservation planning and make no claim that the prospects from existing plans are dismal. We do argue, however, that the efficiency of conservation schemes can be improved by detailing where species across multiple higher taxa are found.

Richard Grenyer¹, C. David L. Orme², T. Jonathan Davies³, Sarah F. Jackson⁴, Richard G. Davies⁵, Valerie A. Olson⁶, Kate E. Jones⁷, Kevin J. Gaston⁴, Tim M. Blackburn⁷, Gavin H. Thomas¹, Peter M. Bennett^{7†}, Pamela C. Rasmussen⁸, Tzung-Su Ding⁹, John L. Gittleman¹⁰ & Ian P. F. Owens^{1,2}

¹NERC Centre for Population Biology, Division of Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK.

e-mail: grenyer@imperial.ac.uk

²Division of Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK.

³National Center for Ecological Analysis and Synthesis, 735 State Street, Suite 300, Santa Barbara, California 93101, USA.

⁴Biodiversity and Macroecology Group, Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK.

⁵Centre for Ecology, Evolution and Conservation, School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

⁶Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK.

⁷Institute of Zoology, Zoological Society of London, Regents Park, London NW1 4RY, UK.

⁸Michigan State University Museum and Department of Zoology, East Lansing, Michigan 48824-1045, USA.

⁹School of Forestry and Resource Conservation, National Taiwan University, Taipei 106, Taiwan.

¹⁰Odum School of Ecology, University of Georgia, Athens, Georgia 30602-2202, USA.

†Present address: Durrell Institute of Conservation and Ecology, Marlowe Building, University of Kent, Canterbury, Kent CT2 7NR, UK.

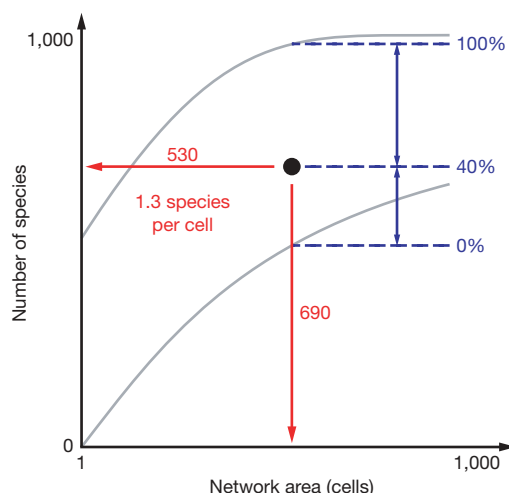


Figure 1 | Alternative performance measures. Rodrigues' preferred measure¹ (blue) evaluates the difference between the maximum possible (upper grey line) and mean random (lower grey line) species capture, given the size of the network in question (black dot). Our preferred measure² (red) evaluates the ratio of species captured to the number of cells in the network.

1. Rodrigues, A. S. L. Effective global conservation strategies. *Nature* 450, doi:10.1038/nature06374 (2007).

2. Grenyer, R. et al. Global distribution and conservation of rare and threatened vertebrates. *Nature* 444, 93–96 (2006).

3. Legendre, L. & Legendre, P. *Numerical Ecology* 2nd edn (Elsevier Science, Amsterdam, 1998).

doi:10.1038/nature06375

Effective global conservation strategies

Arising from: R. Grenyer *et al.* *Nature* **444**, 93–96 (2006).

Using data on the global distribution of mammal¹, bird² and amphibian³ species, Grenyer *et al.*⁴ conclude that planning based on individual taxa does not provide efficient solutions for the conservation of other taxa. They also report that the performance of existing global conservation strategies—endemic bird areas⁵, biodiversity hotspots⁶ and global 200 ecoregions⁷—in representing those taxa is often no better (and in some cases worse) than random. I argue here that the methodology used by Grenyer *et al.* was not appropriate for purported globally comprehensive analyses. Focusing on analyses of rare species as an example, I demonstrate how the data actually reveal substantial cross-taxon surrogacy and good performance of existing global conservation strategies.

Grenyer *et al.* describe the scope of their findings as ‘global’, but their analyses were only performed on subsets of their data. For example, when investigating the congruence between rare birds and rare mammals, they did not obtain a global correlation (across ~19,500 land-grid cells) but a correlation across cells holding rare species of birds or mammals (~5,800 cells). Hence, they excluded ~13,700 cells in which there was perfect agreement between rare birds and rare mammals, in that both were absent. Consequently, the correlation coefficient reported ($r = 0.48$) underestimates the true global correlation ($r = 0.58$). The same applies to all other r values reported, including those in their Fig. 2a–c (ref. 4).

The analyses of species numbers expected by chance in areas of variable size (random curves in their Fig. 4; ref. 4) suffer from the same methodological bias: for instance, the curve in their Fig. 4b is not for a true random selection, but for a random selection across just the ~6,000 cells containing rare species. The true global random curve is much shallower (Fig. 1). Globally, endemic bird areas, hotspots and global 200 ecoregions perform substantially better in

representing rare vertebrate species than would be expected by chance (Fig. 1). Given that hotspots were selected on the basis of plant endemism, this provides evidence for surrogacy of restricted-range plants in representing rare vertebrates.

Grenyer *et al.* investigated cross-taxon surrogacy by counting how many target species are represented in minimum complementary sets selected for a particular surrogate taxon (their Table 1; ref. 4). For rare species, they found values ranging from 22.5% to 77.9% and concluded that surrogacy is low. However, these values alone are not informative: they need to be compared with what would be expected by chance, and what the maximum possible representation is, in an area of the same size⁸. Figure 1 provides this information for when the target is the representation of rare species across the three groups. It shows that minimum sets representing rare mammals, rare birds or rare amphibians individually represent substantially more overall rare species than would be expected by chance. Furthermore, these minimum sets are noticeably close to the maximum representation possible, which is indicative of a high degree of surrogacy.

Analyses of rare species are the most disrupted by the methodological problems described here, but the other two groups analysed by Grenyer *et al.* (all species and threatened species) are also affected.

In conclusion, the analyses in Grenyer *et al.* suffer from a systematic methodological bias that does not allow the results to be compared with the maximal possible representation. The prospects for global conservation planning are, in fact, positive, not dismal as portrayed⁴. It is true that better results will be obtained when high-resolution data become available for all taxa we aim to conserve. Nonetheless, at least for the terrestrial realm, good progress can be, and has already been, achieved by conservation planning based on existing data.

METHODS

The following databases were used: ADHoC database of geographic ranges of birds², owned and developed by the NERC Avian Diversity Hotspots Consortium; global mammal database^{1,4}, owned and developed by J. Gittleman; Global Amphibian Assessment^{3,9}, developed by SSC-IUCN, CABS-CI and NatureServe. Rare species are those in the lower quartile of the range distribution of each taxonomic group⁴. Optimizations were achieved with the GNU Linear Programming Kit package.

Ana S. L. Rodrigues¹

¹Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

e-mail: aslr2@cam.ac.uk

Received 24 July; accepted 25 September 2007.

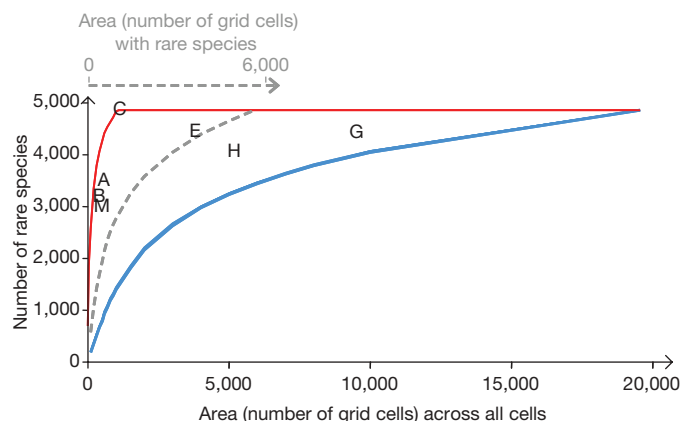


Figure 1 | Relative performance of different types of priority network in capturing rare species (see Fig. 4b of Grenyer *et al.*⁴). Minimum complementary set representing each species at least once of rare mammals (M), rare birds (B), rare amphibians (A), and rare mammals, rare birds and rare amphibians combined (C). Position is also shown for biodiversity hotspots (H), endemic bird areas (E) and global 200 ecoregions (G). Performance is evaluated by comparing the position of these points for equivalent areas with the maximum number of rare species that can be represented (red line) and the number of rare species expected by chance (blue: 95% confidence range for randomly selected sets of cells; 100 replicates). The dashed line is the random line from Grenyer *et al.*⁴, which corresponds to a random selection across just the sites holding rare species, plotted on the dashed grey x axis, top. In Fig. 4 of Grenyer *et al.*⁴, the position of the random lines therefore cannot be compared with the positions of points E, H and G as they are plotted on different axes (the positions of points M, B, A and C are the same on both axes).

1. Sechrest, W. *Global Diversity, Endemism and Conservation of Mammals*. Thesis, Univ. Virginia (2003).
2. Orme, C. D. L. *et al.* Global hotspots of species richness are not congruent with endemism or threat. *Nature* **436**, 1016–1019 (2005).
3. Stuart, S. N. *et al.* Status and trends of amphibian declines and extinctions worldwide. *Science* **306**, 1783–1786 (2004).
4. Grenyer, R. *et al.* Global distribution and conservation of rare and threatened vertebrates. *Nature* **444**, 93–96 (2006).
5. Stattersfield, A. J., Crosby, M. J., Long, A. J. & Wege, D. C. *Endemic Bird Areas of the World: Priorities for Biodiversity Conservation* (BirdLife International, Cambridge, UK, 1998).
6. Mittermeier, R. A., *et al.* *Hotspots Revisited: Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions* (CEMEX, Conservation International and Agrupación Sierra Madre, Mexico City, Mexico, 2004).
7. Olson, D. & Dinerstein, E. The Global 200: Priority ecoregions for global conservation. *Ann. Mo. Bot. Gard.* **89**, 199–224 (2002).
8. Ferrier, S. Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? *Syst. Biol.* **51**, 331–363 (2002).
9. IUCN, Conservation International and NatureServe. *Global Amphibian Assessment* (<http://www.globalamphibians.org>) (2006).

doi:10.1038/nature06374

Grenyer et al. reply

Replying to: A. S. L. Rodrigues *Nature* 450, doi:10.1038/nature06374 (2007).

Rodrigues¹ criticizes our demonstration² of low congruence in the global distributions of rare and threatened vertebrates on the grounds that we excluded locations where species counts were zero from our analyses. In practice, this makes no substantive difference to our conclusions. Some sample locations are not inhabited by any organisms of interest: such locations can inflate measures of covariation and association because their values for parameters of interest (in this case, zero counts of species) are identical. This bias has long been known (as 'the double-zero problem'³); many ecological techniques and studies exclude double-zero data for this reason.

We are happy to clarify that we excluded double-zero data. We concur with Rodrigues¹ that including double-zero data increases the observed correlation coefficient (r). Indeed, it cannot fail to do so: r becomes a function of the ratio of double-zero to non-zero data. For example, given our 19,564 terrestrial grid cells, if fewer than 13,043 cells (two-thirds of the world's surface) are filled with perfectly negatively correlated richnesses and the rest with double zeroes, a positive correlation is observed. In any such case, including double-zero data would be to predetermine the result: our comparison of rare mammals and amphibians involved only 3,184 non-zero cells. This is sufficient justification for our choice of method.

Treating all cells as informative, as Rodrigues proposes, becomes analogous to the analysis of binary presence-absence data: it examines associations between cells that have any species of interest, rather than the relative diversities of those cells. Such questions are better addressed by explicitly converting species counts to binary presence-absence data before analysis. The analyses would rely more on the

veracity of data on absence, and the metrics of association will depend on the relative abundance of zero to non-zero data.

Rodrigues also criticizes our description of network performance; our two contrasting methods are shown in Fig. 1. Our definition equates performance with the trade-off between network size and species capture. Rodrigues prefers a measure that quantifies the effectiveness of a network given its size. Both measures describe different aspects of reserve network performance. We prefer our definition in this case because we regard the comparison of absolute network sizes to be important. For example, under Rodrigues' definition, the optimal network for rare mammals and the global 200 scheme (points M and G in Rodrigues' Fig. 1; ref. 1) have an effectiveness of around 70% and 50%, respectively, despite differing in area by more than an order of magnitude.

We agree that good progress has been made in conservation planning and make no claim that the prospects from existing plans are dismal. We do argue, however, that the efficiency of conservation schemes can be improved by detailing where species across multiple higher taxa are found.

Richard Grenyer¹, C. David L. Orme², T. Jonathan Davies³, Sarah F. Jackson⁴, Richard G. Davies⁵, Valerie A. Olson⁶, Kate E. Jones⁷, Kevin J. Gaston⁴, Tim M. Blackburn⁷, Gavin H. Thomas¹, Peter M. Bennett^{7†}, Pamela C. Rasmussen⁸, Tzung-Su Ding⁹, John L. Gittleman¹⁰ & Ian P. F. Owens^{1,2}

¹NERC Centre for Population Biology, Division of Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK.

e-mail: grenyer@imperial.ac.uk

²Division of Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK.

³National Center for Ecological Analysis and Synthesis, 735 State Street, Suite 300, Santa Barbara, California 93101, USA.

⁴Biodiversity and Macroecology Group, Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK.

⁵Centre for Ecology, Evolution and Conservation, School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

⁶Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK.

⁷Institute of Zoology, Zoological Society of London, Regents Park, London NW1 4RY, UK.

⁸Michigan State University Museum and Department of Zoology, East Lansing, Michigan 48824-1045, USA.

⁹School of Forestry and Resource Conservation, National Taiwan University, Taipei 106, Taiwan.

¹⁰Odum School of Ecology, University of Georgia, Athens, Georgia 30602-2202, USA.

†Present address: Durrell Institute of Conservation and Ecology, Marlowe Building, University of Kent, Canterbury, Kent CT2 7NR, UK.

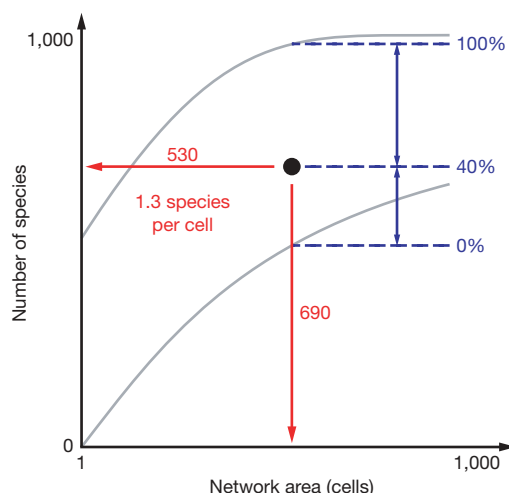


Figure 1 | Alternative performance measures. Rodrigues' preferred measure¹ (blue) evaluates the difference between the maximum possible (upper grey line) and mean random (lower grey line) species capture, given the size of the network in question (black dot). Our preferred measure² (red) evaluates the ratio of species captured to the number of cells in the network.

1. Rodrigues, A. S. L. Effective global conservation strategies. *Nature* 450, doi:10.1038/nature06374 (2007).

2. Grenyer, R. et al. Global distribution and conservation of rare and threatened vertebrates. *Nature* 444, 93–96 (2006).

3. Legendre, L. & Legendre, P. *Numerical Ecology* 2nd edn (Elsevier Science, Amsterdam, 1998).

doi:10.1038/nature06375

Snowball Earth prevention by dissolved organic carbon remineralization

W. Richard Peltier¹, Yonggang Liu¹ & John W. Crowley¹

The ‘snowball Earth’ hypothesis posits the occurrence of a sequence of glaciations in the Earth’s history sufficiently deep that photosynthetic activity was essentially arrested. Because the time interval during which these events are believed to have occurred immediately preceded the Cambrian explosion of life, the issue as to whether such snowball states actually developed has important implications for our understanding of evolutionary biology. Here we couple an explicit model of the Neoproterozoic carbon cycle to a model of the physical climate system. We show that the drawdown of atmospheric oxygen into the ocean, as surface temperatures decline, operates so as to increase the rate of remineralization of a massive pool of dissolved organic carbon. This leads directly to an increase of atmospheric carbon dioxide, enhanced greenhouse warming of the surface of the Earth, and the prevention of a snowball state.

During the Neoproterozoic era of the Earth’s history, the carbon cycle exhibited a sequence of oscillation-like variations that included decreases of $\delta^{13}\text{C}$ to levels that have been interpreted to imply the occurrence of intense snowball glaciations¹ during which photosynthetic activity essentially ceased^{2–4}. Such events would have strongly affected the evolution of eukaryotic life and this has led to suggestions that this extreme interpretation of the observed variability in the carbon cycle could be unwarranted^{5,6}. We have developed a coupled model of the co-evolution of Neoproterozoic climate and the carbon cycle that provides an alternative interpretation to the ‘hard snowball’ hypothesis of the origin of the observed $\delta^{13}\text{C}$ variations. This model links a previously developed model of the Neoproterozoic physical climate system^{7,8} to a recently developed model of the carbon cycle⁹ for the same time interval. The coupled model is shown to support a limit cycle oscillation in which the temperature dependence of the solubility of oxygen in sea water controls the rate of remineralization of organic carbon such that the level of atmospheric CO_2 is prevented from becoming sufficiently low to allow a hard snowball state to occur. The model also satisfies the timescale and continental ice volume constraints that have been inferred to characterize these glaciation events, as well as the magnitude of the carbon cycle excursions when an appropriate biogeochemical dependence of photosynthetic carbon isotopic fractionation is assumed.

The sequence of intense glaciations that occurred during the Cryogenian period of the Neoproterozoic era—a period that began approximately 850 million years (Myr) ago and which ended approximately 635 Myr ago with the onset of the Ediacaran period—is currently an intense focus of interdisciplinary activity¹⁰. Evolutionary biologists and palaeontologists^{11,12} are interested in this era because it preceded the Cambrian explosion of life, during which eukaryotic biological diversity proliferated. Climate dynamicists^{13–18} have been attracted by the challenge posed by the appearance in the geological record of continental-scale glaciation that is suggested to have reached sea level in equatorial latitudes at some locations. Sedimentologists^{19–21} and geochronologists²² have worked on the global-scale correlation of glacial units across the present-day continents, during a time of intense tectonic activity involving the break-up of the supercontinent of Rodinia. However, many important

questions are still unresolved. In particular, how many glaciations actually occurred during the Cryogenian period is currently uncertain. Figure 1, a revised and extended version of previously published sketches^{23–25} of the evolution of $\delta^{13}\text{C}$ over the most recent billion years of the Earth’s history, illustrates the connection in time between this measure of climate variability, major tectonic events, and periods of intense glaciation.

The nature of Cryogenian glacial episodes

Although some recent work has championed the notion that only three major glaciations occurred during this period²³—the Sturtian glaciation at 723^{+16}_{-10} Myr ago, the Marinoan glaciation between 659 and 637 Myr ago, and the Gaskiers glaciation at approximately 582 Myr ago—there remain significant issues concerning the synchronicity of these events as inferred on the basis of the stratigraphic record from different continents. The chronological control upon the Sturtian glaciation, in particular, now suggests that it consisted of at least two distinct glacial episodes²². Evidence from the Huqf Supergroup of Oman has recently been interpreted to imply that no hard snowball glaciation event could have occurred^{19–21}. Similarly, the duration of individual glacial episodes remains unknown (although it has been speculated to be between 4 and 30 million years⁴), as is the answer to the question of whether each of the glacial episodes consisted of a single ice advance and retreat or of multiple such events.

A previously proposed model of Neoproterozoic climate suggested the plausibility of a “slushball” solution in which a significant region of open water could have persisted at the Equator during each of these events. This model has been criticized²⁶ on the basis of the claim that it could not explain the inferred requirement of the geological record for the occurrence of episodes of glaciation that lasted at least 4 Myr. This is the point of departure for the analyses described in this paper. Our target has been the question posed by the carbon isotopic variability depicted qualitatively in Fig. 1. Whereas subsequent to the Neoproterozoic era the carbon cycle appears to have been operating in a quasi-equilibrium mode⁹, during the Neoproterozoic it appears to have been operating in an out-of-equilibrium mode that is peculiar to this interval of the Earth’s history. This mode of behaviour was recently suggested⁹ to have arisen as a consequence of a significant

¹Department of Physics, University of Toronto, Toronto, Ontario M5S 1A7, Canada.

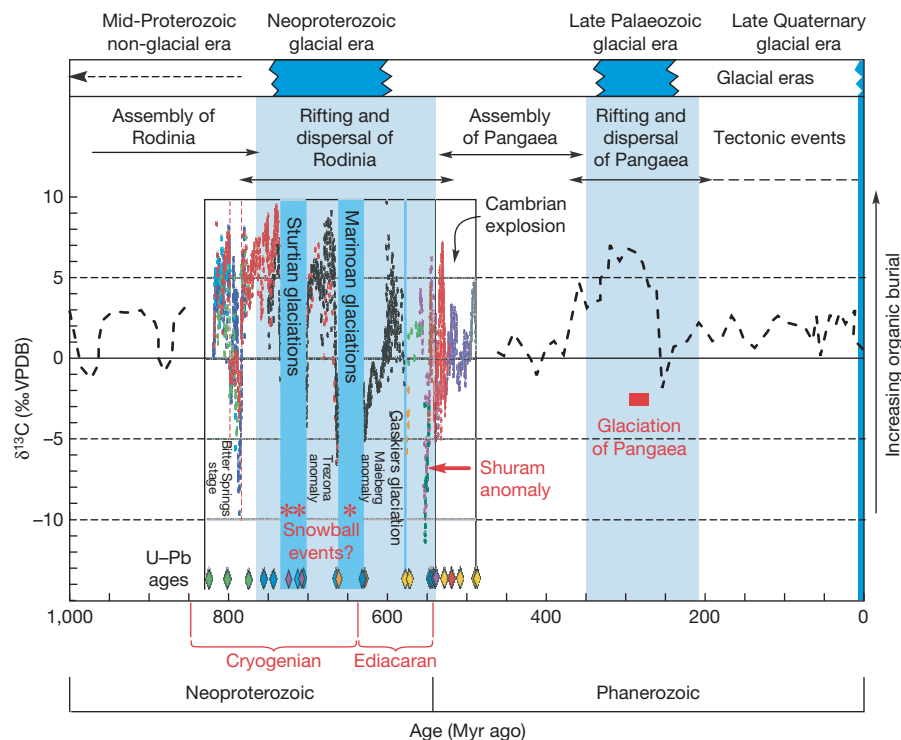


Figure 1 | The history of $\delta^{13}\text{C}_{\text{inorg}}$ variations over the past billion years as measured in sequences of carbonate rocks on land. This quantity provides an indirect measure of photosynthetic activity. High values correspond to times of high rates of burial of organic matter, a consequence of the fact that isotopically light carbon (^{12}C) is preferentially partitioned into organic matter by photosynthesis. Burial of this matter leaves behind an ocean that is isotopically heavier than average, from which the carbonates are precipitated. The Neoproterozoic era was clearly characterized by extreme variability in $\delta^{13}\text{C}_{\text{inorg}}$. The largest-amplitude negative anomaly that is inferred to have existed during the Neoproterozoic is the Shuram anomaly³⁵, which is taken to correspond to the major oxidation event that eliminated the condition

imbalance between the mass of carbon stored in the Neoproterozoic ocean in the organic and inorganic forms. Atmospheric oxygen could have provided an important link between the carbon cycle and climate but no detailed discussion was provided⁹. The purpose of the analyses reported here is to elaborate a plausible linkage and to investigate the extent to which this may shed light upon the state of the Earth's physical climate system during this critical period for biological evolution.

Coupled carbon cycle–climate evolution

A schematic of the model we have developed for this purpose is shown in Fig. 2, consisting of three primary elements: a model of the carbon cycle⁹, a model of the physical climate system (consisting of surface energy balance and sea-ice components⁸), and a detailed model of continental-scale glaciation (the University of Toronto Glacial Systems Model; ref. 27). The mathematical details of the carbon cycle component of the complete coupled model, which involve a significant extension of the model of ref. 9, are described in the Supplementary Information). To link these models we explicitly incorporated a dependence of the remineralization flux J_{21} (Fig. 2), through which organic carbon is converted to inorganic carbon, upon the (temperature-dependent) solubility of oxygen in sea water:

$$J_{21} = J_{21e} [1 - F_{21}(T - T_e)] \quad (1)$$

In this expression, T is the mean surface temperature of the planet determined by the physical climate model and T_e is the equilibrium temperature at which the remineralization flux J_{21} equals its equilibrium value. For a discussion of the physical significance of the

required for out-of-equilibrium behaviour of the carbon cycle, namely the extreme imbalance between the masses of the inorganic and organic reservoirs. This figure is a modified version of those originally produced by Kaufman² (Fig. 2), Eyles and Januszczak²⁵ (Fig. 9) and Halverson *et al.*²³ (observational data used with permission). The data presented through the Cryogenian interval are based upon the recently compiled composite record of ref. 23 and are reproduced here with permission. The diamond-shaped symbols represent the times for which high-quality radiometric age determinations (based on the U–Pb dating of detrital zircons) are available to provide chronological control on the sequence of Neoproterozoic events. VPDB, Vienna Pee-Dee belemnite standard.

control parameter F_{21} , please see Methods. A primary assumption of the present version of the model is that the variations in

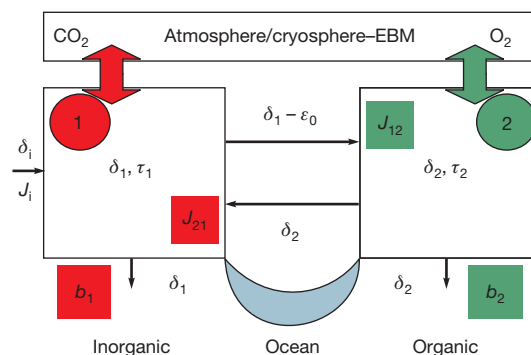


Figure 2 | Diagram of our carbon-cycle-coupled climate model. Its primary component consists of the box model of the carbon cycle of Rothman *et al.*⁹, which is coupled to both an energy balance model of the atmosphere (EBM) and the University of Toronto Glacial Systems Model of continental-scale ice-sheet evolution. The variables $\delta_{1,2}$ and $\tau_{1,2}$ are the isotopic compositions and residence times for the inorganic and organic carbon reservoirs, respectively. δ_i is the input into the inorganic reservoir from, for example, volcanic outgassing (of flux J_i) and ϵ_0 is the isotopic fractionation that occurs in photosynthesis during which inorganic carbon is converted to organic carbon. Fluxes into and out of the two reservoirs are represented by the arrows. Output from the two reservoirs occurs directly by burial (associated with fluxes $b_{1,2}$) as well as by exchanges between them. Crucial to the success of our coupled model of carbon-cycle-climate evolution are the fluxes of O_2 and CO_2 between the atmosphere and ocean components of the model.

atmospheric oxygen that accompany system evolution will have no significant impact upon system dynamics.

The reason that the temperature dependence of the remineralization flux has important implications for the surface climate regime is connected to the fact that as the climate cools and the rate of conversion of organic carbon to inorganic carbon increases, the partial pressure of carbon dioxide in the atmosphere also increases, according to the relation:

$$\frac{p_{\text{CO}_2}(t)}{p_{\text{CO}_2,e}} = \left[\frac{M_1(t)}{M_{1e}} \right]^X \quad (2)$$

in which $p_{\text{CO}_2,e} = 300$ p.p.m.v. is the equilibrium concentration of atmospheric carbon dioxide at temperature T_e , M_1 is the mass of inorganic carbon in the ocean and M_{1e} is the equilibrium mass. It has been suggested²⁸ that the parameter X in equation (2) should be equal to 2 in Phanerozoic circumstances in which the climate is not changing too quickly. It is unclear whether the partition of carbon dioxide between the ocean and the atmosphere during the rapidly changing Neoproterozoic should obey this relationship, so we consider X to be a parameter of the model. As the level of atmospheric CO_2 increases in response to increasing $M_1(t)$, the surface of the planet will be increasingly heated by the increasing infrared flux of energy d_{rad} (in units of W m^{-2}) that is due to 'greenhouse' warming^{29,30}:

$$d_{\text{rad}} = 6.0 \ln \left(\frac{p_{\text{CO}_2}(t)}{p_{\text{CO}_2,e}} \right) \quad (3)$$

The sequence of relationships (1) to (3) together describe a negative feedback process whereby the carbon cycle reacts to a tendency of the planet to cool by enhancing the atmospheric concentration of carbon dioxide and thus inhibiting the cooling. During warming the same feedback operates so as to inhibit this tendency as well. The new coupled model of carbon cycle–climate evolution that we have developed for the Neoproterozoic is strongly controlled by this feedback process.

Cyclic glaciation due to carbon cycle coupling

In the absence of explicit coupling to the carbon cycle, the ice-sheet-coupled energy-balance model of the process of global glaciation

produces steady-state solutions for the mean surface temperature that are a strong function of the concentration of carbon dioxide in the atmosphere, of the solar constant, and of the spatial distribution of the continents. Although the detailed palaeogeography of the Neoproterozoic varied appreciably during the break-up of Rodinia, there is general agreement that during the Sturtian episode(s) the continental fragments were clustered around the Equator. During the Marinoan episode, however, the equatorial positioning of the main land masses was apparently less pronounced. Here we use the same palaeogeography as was used in a previous analysis⁷ of Neoproterozoic surface temperature conditions, one that is more appropriate to the Marinoan glaciation than to the Sturtian(s). For the value of the solar constant we assume a decrease of 6% below the present value, as is appropriate to this stage of the evolution of the Sun.

In the absence of carbon cycle coupling, Fig. 3a shows the steady-state variations of mean surface temperature predicted by the ice-sheet-coupled EBM, as a function of atmospheric carbon dioxide concentration. This model exhibits hysteresis, such that a range of values of the parameter d_{rad} , and thus p_{CO_2} , exists within which at least two different steady states are equally acceptable solutions. Which state is physically realized depends upon the initial conditions of integration from which the solution is approached, as indicated by the arrows on the different branches of the diagram of steady-state solutions. The hysteresis diagram defines a 'hot branch' of solutions as well as an 'oasis branch', the latter being the branch on which the 'slushball' solutions to the problem of Neoproterozoic climate were originally discovered⁸. In this model the hard snowball regime also exists, but to reach it requires that $p_{\text{CO}_2}(t)$ reaches sufficiently low values. To escape from the hard snowball state in this model requires approximately 0.3 bar of atmospheric carbon dioxide if the sea-ice albedo is assumed to be equal to 0.6 in the hard snowball regime⁷. This is somewhat lower than a more accurate estimate later obtained with a more complex model³¹. Here we investigate how the climate model will behave when coupled to the explicit model of the carbon cycle.

The results delivered by this model in fully coupled synchronous mode are illustrated in Figs 3b–e, for which examples we have assumed $X = 2$ in equation (2). Figure 3b shows several cycles of

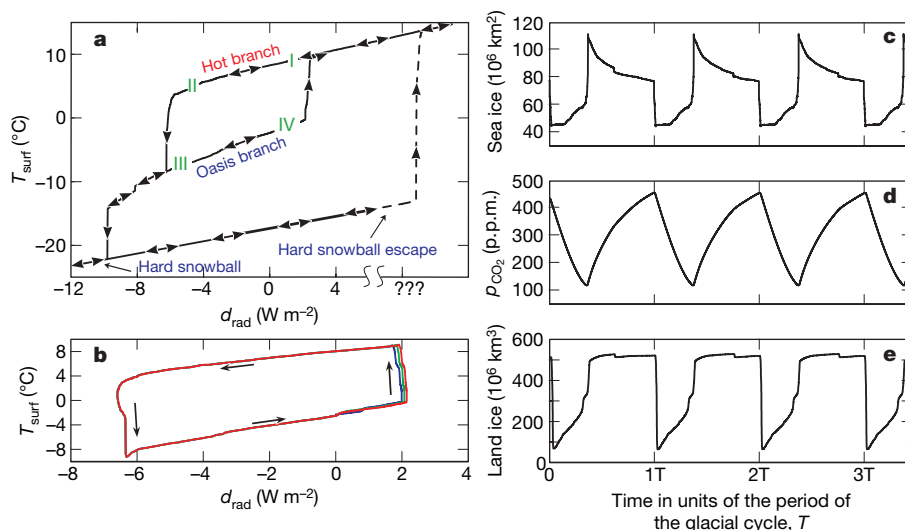


Figure 3 | The cyclic glaciation dynamics of the coupled climate-carbon cycle model. **a**, The steady-state (equilibrium) solutions of the energy-balance-coupled ice-sheet model are shown as a function of the atmospheric carbon dioxide concentration, represented by the increase or decrease of the infrared flux, d_{rad} , received at the surface. The hysteresis in the model state space is indicated by the multiple equilibria for a wide range of values of the carbon dioxide concentration. **b**, The trajectory of solutions of the coupled climate-carbon cycle model in the space of mean surface temperature versus

surface infrared forcing. The solution simply cycles the hysteresis loop of the steady-state solutions when the impact of the temperature dependence of the solubility of oxygen upon the remineralization flux is introduced into the dynamical system. **c–e**, Time series for sea-ice area, atmospheric carbon dioxide concentration and land-ice volume are shown, respectively, for several circuits of the hysteresis loop. The scale of the x axis depends on the control parameter F_{21} , as demonstrated in Fig. 4.

the 'orbit' of the solution in the plane of mean surface temperature versus d_{rad} . These results demonstrate that the coupled system evolves in such a way as to continuously cycle the hysteresis loop defined by the set of steady-state solutions. In Figs 3c–e we show time series for sea-ice area, atmospheric CO_2 and continental ice volume respectively, for several circuits of the hysteresis loop. These results demonstrate that system evolution is governed by a limit cycle oscillation (a periodic solution of the nonlinear dynamical system) which would have a period of approximately 4 Myr for $F_{21} = 8 \times 10^{-5}$ and a period of approximately 30 Myr for $F_{21} = 1 \times 10^{-5}$. Detailed analysis (not shown) demonstrates that the onset of the periodic behaviour for $X = 1$ occurs for $F_{21} < F_{21, \text{crit}}$, where the critical value is in the range $5 \times 10^{-4} < F_{21, \text{crit}} < 6 \times 10^{-4}$. For values less than this, the period of the limit cycle oscillation continues to increase as the value of the control variable decreases. For $X = 2$, the critical value of F_{21} is shifted to the range $3 \times 10^{-4} < F_{21, \text{crit}} < 4 \times 10^{-4}$.

Figure 4, based upon analyses for several values of the control parameter, for which time series are shown in Supplementary Fig. 1 (for $X = 1$) and Supplementary Fig. 2 (for $X = 2$), shows that the period of the limit cycle covers the previously inferred range of durations of the Neoproterozoic glacial intervals from 4 to ~30 Myr ago. In the limit $F_{21} \rightarrow 0$, the previous model⁹ is recovered and the system exhibits no dynamics, as physical climate is now uncoupled from the carbon cycle. For $F_{21} > F_{21, \text{crit}}$ the system exits the limit cycle regime (see Supplementary Fig. 1 for $F_{21} = 6 \times 10^{-4}$) and a new steady-state solution is realized, one for which the mean surface temperature is equal to the value T_c (not shown). We interpret this 'fixed point' to represent the in-equilibrium solution for the carbon cycle that has apparently been characteristic for the Phanerozoic eon that followed the Neoproterozoic with the onset of the Cambrian. The model therefore embodies a potential explanation not only of the apparently out-of-equilibrium behaviour characteristic of the Neoproterozoic, but also of the in-equilibrium behaviour that has thereafter generally been the rule⁹. Supplementary Fig. 3 shows normalized time series for all model fields as a means of illustrating the phase relationships between them.

Eustatic sea level and ice volume variations

Figure 5 shows the distribution of land ice and sea ice over the surface of the Earth that the model delivers at each of the corners (labelled I to IV on Fig. 3a) of the hysteresis loop of steady solutions onto which the solution is entrained in the time-dependent limit cycle regime. This regime is essentially perfectly periodic, so this set of four solutions fully characterizes the extremes of system behaviour. The low-temperature,

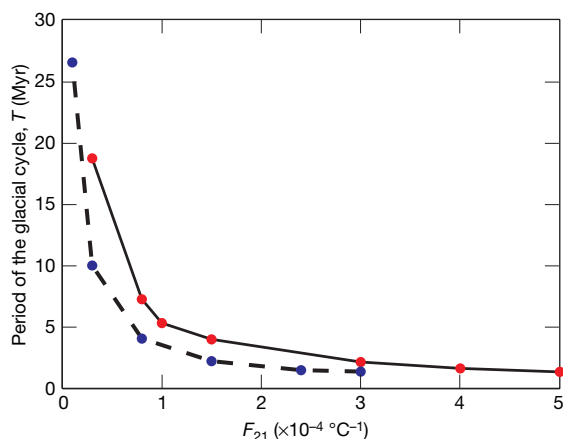


Figure 4 | The period of the glacial cycle predicted by the coupled climate–carbon cycle model as a function of the control parameter of the model F_{21} . Results are shown for both $X = 1$ (solid line) and $X = 2$ (dashed line), where X determines how the atmospheric carbon dioxide concentration changes as the mass of inorganic carbon changes relative to the reference values (equation (2)). The predicted period spans the range previously inferred to have been characteristic of Neoproterozoic glacial intervals.

low-carbon dioxide corner of the loop (located at III in Fig. 3a) shows that although most of the continental fragments are covered by a thick veneer of land ice and much of the surface of the ocean is covered by sea ice, there is still a substantial area of the equatorial ocean that consists of open water. This is the 'slushball' regime discovered previously⁸. The maximum volume of land ice that develops during a single glacial cycle is approximately $500 \times 10^6 \text{ km}^3$, as shown on Fig. 3e, which would correspond to a fall of eustatic sea level of approximately 1 km, a result that could be interpreted as consistent with the deep Neoproterozoic canyons that have been investigated on the Australian continent^{32,33}. Although much larger, this sea-level depression is still reasonable compared to the sea-level drop of about 120 m that occurred during the Last Glacial Maximum³⁴, when polar ice sheets covered only a small fraction of the land surface.

We therefore interpret the Cryogenian period as a time during which, in an organic-carbon-rich environment, the mass of organic matter in the Neoproterozoic ocean had reached such a large value relative to the mass of inorganic carbon that out-of-equilibrium behaviour with strong coupling to the physical climate system became possible. In our interpretation, this coupling ceased to be possible after the intense oxidation of the organic reservoir that occurred towards the end of the Ediacaran period and just before the onset of the Cambrian²². We interpret this oxidation event as coinciding with the Shuram anomaly shown on Fig. 1³⁵.

The isotopic variability of inorganic carbon

The model also predicts the time series of $\delta^{13}\text{C}(t)$ for the inorganic reservoir, the primary diagnostic that has been taken to require the existence of the hard snowball regime. Escape from this regime is assumed to have been possible only via the build-up of the atmospheric carbon dioxide concentration owing to the presumed slow and continuous action of volcanic degassing while the oceans were entirely covered by sea ice. To make contact with the different isotopic concentrations of organic and inorganic carbon that develop during a single glacial cycle, the isotopic fractionation ε_0 that occurs during photosynthesis (see Fig. 2) must also depend upon ocean biogeochemistry, and thus climate. To predict the evolution of carbon isotopes we may therefore assume:

$$\varepsilon_0 = \varepsilon_e + \beta_{\text{frac}, M} (M_1 [\text{CO}_2(t)] - M_{1c}) \quad (4)$$

This representation of the isotopic fractionation that occurs during photosynthesis is intended to provide a parameterization of the

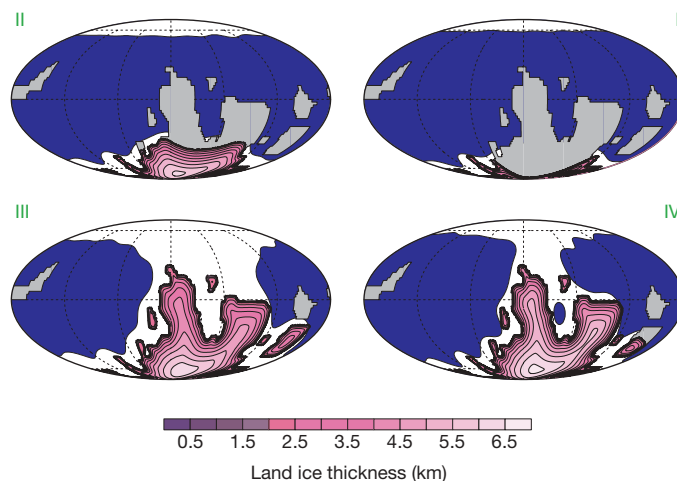


Figure 5 | Examples of the distribution of land ice and sea ice during the glacial cycle. These examples correspond to the climate states occupied by the model at each of the four corners (labelled I to IV) of the hysteresis loop of steady solutions shown on Fig. 3a. Irrespective of the value of the control parameter, in the coupled mode the dynamical system simply continues to cycle the hysteresis loop. Land-ice thickness is contoured in shades of mauve, and sea ice is shown as white.

influence of kinetic isotope effects associated, say, with the enzymatically catalysed carboxylation of ribose biphosphate (J. M. Hayes, personal communication, 2006). The mass of the inorganic reservoir M_1 is expected to provide a reasonably direct representation of the main control on this catalytic process. This simple parameterization appears to be in acceptable accord with the analyses of ref. 36 as used in the discussion²⁸ of the dependence of the isotopic difference between organic and inorganic carbon as a function of atmospheric p_{CO_2} . Best fits of the model prediction of the time variation of $\delta^{13}\text{C}(t)$ through a typical Neoproterozoic glacial–interglacial cycle based upon the use of equation (4) are obtained with a value of $\beta_{\text{frac},M} \approx 0.00048\%$ per gigaton. This value is not unreasonable given that the species composition of the hypothesized organic-carbon-heavy Neoproterozoic ocean, in which $M_2 \gg M_1$, is largely unknown.

Figure 6a for $X = 2$ shows the model prediction of the evolution of $\delta^{13}\text{C}_{\text{inorg}}$ as a function of $\varepsilon = \delta_1 - \delta_2$, through a single Neoproterozoic glacial–interglacial cycle, compared with the complete data set from ref. 9, inspection of which demonstrates the good fit that the model is capable of delivering to the data. When the comparison is limited to the interval of time from 738 to 593 Myr ago that contains the Sturtian and Marinoan events (Fig. 6b), however, the fit of the model to the data significantly improves. This demonstrates that the carbon isotopic measurements from the interval containing the hypothesized hard snowball events are well explained by the carbon-cycle-coupled slushball model. As discussed in ref. 9, the slope of the model fit to the data in the $\delta^{13}\text{C}_{\text{inorg}}$ versus ε plane is unity, which is highly diagnostic of the condition $M_2 \gg M_1$ that is characteristic of the carbon cycle component of the model. Supplementary Fig. 4 shows equivalent results for the case $X = 1$.

Future model developments

Among several further avenues of investigation as we continue to develop this model of the coupled evolution of the carbon cycle and climate during the Neoproterozoic, one we intend to explore concerns the potential of the model as an explicit means of assessing the magnitude of the variations of atmospheric oxygen that would be expected to accompany the glaciation–deglaciation process. Another concerns whether oxidants other than oxygen, such as sulphate, may have an important role to play. Also, how might the drawdown of atmospheric carbon dioxide from the atmosphere due to the weathering of calcium and magnesium silicates be expected to affect our model predictions¹⁶? Here we neglected this drawdown because of the strong temperature and precipitation dependence that is characteristic of the silicate weathering process, such that low temperature and precipitation and the absence of vascular plants in this period of the Earth's history would have significantly reduced its effect^{37–39}. In spite of this and other remaining uncertainties, the new model strongly suggests that those

observations that have been assumed to be most diagnostic of hard snowball conditions may not require such an extreme interpretation. In our view it is more likely that it may have been the carbon cycle, not the physical climate system, that was operating in an extreme mode before the onset of the Cambrian explosion of life.

METHODS SUMMARY

Our model consists of three primary components. (1) A global surface energy balance model in spherical geometry with a specified distribution of continents and ocean coupled to a simple thermodynamic model of sea-ice formation. This element of the structure is time-dependent and may be subjected to orbital insolation variations. It also resolves the annual cycle of surface temperature. (2) A detailed model of the growth and evolution of continental ice sheets that occurs in response to variations in the surface temperature and precipitation regimes. In this model, precipitation forcing of the hydrological cycle is prescribed. Furthermore, the glacial isostatic adjustment process that affects the flow of ice over the surface of a continent under the action of the gravitational force is fully incorporated. (3) A two-box model of the ocean carbon cycle that differentiates between organic and inorganic reservoirs. Components (1) and (2) are described in terms of coupled nonlinear partial differential equations whereas the third component is described in terms of coupled nonlinear ordinary differential equations. On the long timescales that are of interest to us we may safely assume the concentration of carbon dioxide to be uniform throughout the atmosphere at all times, because CO_2 is a well-mixed atmospheric trace gas. The three model components are coupled together through the action of an assumed temperature dependence of the remineralization flux that converts organic carbon to its inorganic (CO_2) form. The temperature assumed to control the remineralization flux is taken to be the mean surface temperature of the planet, a temperature that is dominated by the sea-surface temperature of the tropical ocean.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 31 October 2006; accepted 1 October 2007.

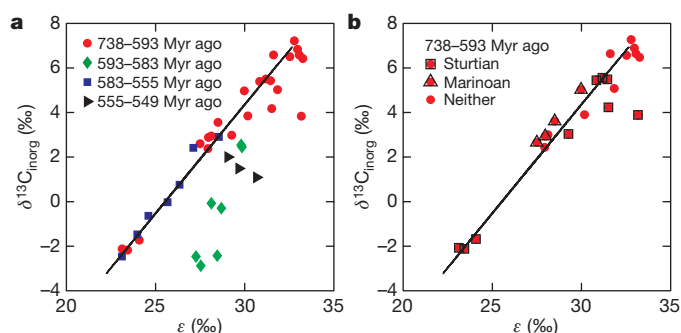


Figure 6 | Model predictions of $\delta^{13}\text{C}_{\text{inorg}}(t)$ as a function of $\varepsilon(t)$ through a single glacial cycle. These examples for the case $X = 2$ are compared with observations of contemporaneous samples of the isotopic ratios of organic and inorganic carbon at a large number of times through the Neoproterozoic. **a**, Comparison of all of the available data from Rothman *et al.*⁹. **b**, The data derived from 738–593 Myr ago are shown separately and the data points from the Sturtian and Marinoan intervals are singled out.

- Kirschvink, J. L. In *The Proterozoic Biosphere: A Multi-Disciplinary Study* (eds Schopf, J. W. & Klein, C.) 51–52 (Cambridge Univ. Press, Cambridge, UK, 1992).
- Kaufman, A. J. An ice age in the tropics. *Nature* **386**, 227–228 (1997).
- Hoffman, P. F., Kaufman, A. J., Halverson, G. P. & Schrag, D. P. A Neoproterozoic Snowball Earth. *Science* **281**, 1342–1346 (1998).
- Hoffman, P. F. & Schrag, D. P. The snowball Earth hypothesis: testing the limits of global change. *Terra Nova* **14**, 129–155 (2002).
- Runnegar, B. Loophole for snowball Earth. *Nature* **405**, 403–404 (2000).
- Knoll, A. H. *Life on a Young Planet: The First Three Billion Years of Evolution on Earth* 1–277 (Princeton Univ. Press, Princeton, New Jersey, 2003).
- Hyde, W. T., Crowley, T. J., Baum, S. K. & Peltier, W. R. Neoproterozoic “Snowball Earth” simulations with a coupled climate/ice-sheet model. *Nature* **405**, 425–429 (2000).
- Peltier, W. R., Tarasov, L., Vettoretti, G. & Solheim, L. P. In *The Extreme Proterozoic: Geology, Geochemistry and Climate* (eds Jenkins, G. S. *et al.*) *Geophys. Monogr. Ser.*, **146**, 107–124 (AGU Press, Washington DC, 2004).
- Rothman, D. H., Hayes, J. M. & Summons, R. E. Dynamics of the Neoproterozoic carbon cycle. *Proc. Natl Acad. Sci. USA* **100**, 8124–8129 (2003).
- Allen, P. A. Snowball Earth on trial. *Eos* **87**, 495 (2006).
- Moczydlowska, M. Neoproterozoic and Cambrian successions deposited on the east European platform and Cadomian basement area of Poland. *Stud. Geophys. Geodactica* **39**, 276–285 (1995).
- Grey, K. & Corkran, M. Late Neoproterozoic stromatolites in glacial successions of the Kimberley region, Western Australia: evidence for a younger Marinoan glaciation. *Precamb. Res.* **92**, 65–87 (1998).
- Chandler, M. A. & Sohl, L. E. Climate forcings and the initiation of low-latitude ice sheets during the Neoproterozoic Varanger glacial interval. *J. Geophys. Res.* **105**, 20737–20756 (2000).
- Goodman, J. C. & Pierrehumbert, R. T. Glacial flow of floating marine ice in “Snowball Earth”. *J. Geophys. Res.* **108**, doi:10.1029/2002JC001471 (2003).
- Poulsen, C. J. & Jacob, R. L. Factors that inhibit snowball Earth simulation. *Paleoceanography* **19**, doi:10.1029/2004PA001056 (2004).
- Donnadieu, Y., Godd  ris, Y., Ramstein, G., N  d  lec, A. & Meert, J. A. “Snowball Earth” triggered by continental breakup through changes in runoff. *Nature* **428**, 303–306 (2004).
- Pollard, D. & Kasting, J. F. Snowball Earth: a thin ice solution with flowing sea glaciers. *J. Geophys. Res.* **110**, doi:10.1029/2004JC002525 (2005).
- Pierrehumbert, R. T. Climate dynamics of a hard snowball Earth. *J. Geophys. Res.* **110**, doi:10.1029/2004JD005162 (2005).
- Leather, J., Allen, P. A., Brasier, M. D. & Cozzi, A. Neoproterozoic snowball Earth under scrutiny: Evidence from the Fiq glaciation of Oman. *Geology* **30**, 891–894 (2002).

20. Le Guerroué, E., Allen, P. A. & Cozzi, A. Chemostratigraphic and sedimentological framework of the largest negative carbon isotopic excursion in Earth history: The Neoproterozoic Shuram formation (Nafun Group, Oman). *Precamb. Res.* **146**, 68–92 (2006).
21. Allen, P. A., Leather, J. & Brasier, M. D. The Neoproterozoic Fiq glaciation and its aftermath: Huqf Supergroup of Oman. *Basin Res.* **16**, 507–534 (2004).
22. Condon, D. *et al.* U–Pb ages from the Neoproterozoic Doushantuo formation, China. *Science* **308**, 95–98 (2005).
23. Halverson, G. P., Hoffman, P. F., Schrag, D. P., Maloof, A. C. & Rice, A. H. N. Towards a Neoproterozoic composite carbon cycle isotopic record. *Geol. Soc. Am. Bull.* **117**, 1181–1207 (2005).
24. Hoffman, P. F. & Schrag, D. P. Snowball Earth. *Sci. Am.* **282**, 68–75 (2000).
25. Eyles, N. & Januszczak, N. “Zipper-Rift”: A tectonic model for Neoproterozoic glaciations during the breakup of Rodinia after 750 Ma. *Earth Sci. Rev.* **65**, 1–73 (2004).
26. Schrag, D. P. & Hoffman, P. F. Life, geology, and snowball Earth. *Nature* **9**, 306 (2001).
27. Tarasov, L. & Peltier, W. R. Arctic freshwater forcing of the Younger-Dryas cold reversal. *Nature* **435**, 662–665 (2005).
28. Kump, L. R. & Arthur, M. A. Interpreting carbon-isotope excursions: carbonates and organic matter. *Chem. Geol.* **161**, 181–198 (1999).
29. Ramanathan, V., Lian, M. S. & Cess, R. D. Increasing atmospheric CO₂: zonal and seasonal estimates of the effect on the radiation energy balance and surface temperature. *J. Geophys. Res.* **84**, 4949–4958 (1979).
30. Marshall, H. G., Walker, J. C. G. & Kuhn, W. R. Long-term climate change and the geochemical cycle of carbon. *J. Geophys. Res.* **93**, 791–801 (1988).
31. Pierrehumbert, R. T. High levels of atmospheric carbon dioxide necessary for the termination of global glaciation. *Nature* **429**, 646–649 (2004).
32. Christie-Blick, N., von der Borch, C. C. & Dibona, P. A. Working hypotheses for the origin of the Wonoka canyons (Neoproterozoic), South Australia. *Am. J. Sci.* **A 290**, 295–332 (1990).
33. Christie-Blick, N., Williams, G. E. & Gostin, V. A. Discussion on mantle plume uplift in the sedimentary record: Origin of kilometre-deep canyons within late Neoproterozoic successions, South Australia. *J. Geol. Soc.* **158**, 573–576 (2001).
34. Peltier, W. R. & Fairbanks, R. G. Global glacial ice volume and Last Glacial Maximum duration from an extended Barbados sea level record. *Quat. Sci. Rev.* **25**, 3322–3337 (2006).
35. Fike, D. A., Grotzinger, J. P., Pratt, L. M. & Summons, R. E. Oxidation of the Ediacaran ocean. *Nature* **444**, 744–747 (2006).
36. Bidigare, R. R. *et al.* Consistent fractionation of ¹³C in nature and in the laboratory: growth-rate effects in some haptophyte algae. *Glob. Biogeochem. Cycles* **11**, 279–292 (1997).
37. Walker, J. C. G., Hays, P. B. & Kasting, J. F. A negative feedback mechanism for the long term stabilization of earth's surface temperature. *J. Geophys. Res.* **86**, 9776–9782 (1981).
38. Berner, R. A., Lasaga, A. C. & Garrels, R. M. The carbonate-silicate geochemical cycle and its effect on atmospheric carbon dioxide over the past 100 million years. *Am. J. Sci.* **283**, 641–683 (1983).
39. Berner, R. A. *The Phanerozoic Carbon Cycle: CO₂ and O₂ 1–150* (Oxford Univ. Press, Oxford, UK, 2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This paper is a contribution to the Polar Climate Stability Network, which is sponsored by the Canadian Foundation for Climate and Atmospheric Science and a consortium of Canadian universities. Additional assistance was provided by the Natural Sciences and Engineering Research Council of Canada.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to W.R.P. (peltier@atmos.physics.utoronto.ca).

METHODS

By far the most important aspect of the coupled carbon cycle–climate model that we have developed concerns the feedback between temperature and the rate at which organic carbon is remineralized to produce inorganic carbon, as expressed in equation (1). This relationship is supported by the well-known dependence, through Henry's law of classical thermodynamics, of the solubility of a gas in a liquid upon the temperature of the liquid. Garcia *et al.*⁴⁰ provide detailed information on the basis of which we justify a linear parameterization of the solubility of oxygen in sea water:

$$O_{2, \text{sol}} = O_{2, \text{sol}_e} - A(T - T_e) \quad (5)$$

This dependence is introduced into the carbon cycle model by assuming that the remineralization flux J_{21} is itself linearly dependent upon the deviation of the solubility of oxygen away from its value at an equilibrium temperature T_e . An appropriate modification of the equilibrium flux J_{21_e} that incorporates this dependence (see Supplementary Information) is:

$$J_{21} = J_{21_e} \left\{ 1 + B \left(\frac{O_{2, \text{sol}} - O_{2, \text{sol}_e}}{O_{2, \text{sol}_e}} \right) \right\} \quad (6)$$

Incorporation of equation (5) into equation (6) then gives equation (1) in the main text.

The primary control parameter in the new coupled carbon cycle–climate model is therefore the parameter F_{21} , which has the dimensions of inverse temperature:

$$F_{21} = \frac{AB}{O_{2, \text{sol}_e}} \quad (7)$$

In the low-temperature vicinity of 1 °C, $A \approx 8 \mu\text{mol kg}^{-1} \text{ } ^\circ\text{C}^{-1}$ (ref. 40) and $O_{2, \text{sol}} \approx 400 \mu\text{mol kg}^{-1}$ (ref. 40). Thus $F_{21} \approx 3 \times 10^{-4} \text{ } ^\circ\text{C}^{-1}$ for $B = 1.5 \times 10^{-2}$ (the importance of this value of F_{21} is discussed in the main text). The strength of the dependence of the remineralization flux upon oxygen solubility that is determined by the parameter B is clearly affected by ocean dynamics, including the strength of the overturning circulation that operates under glacial conditions. Although some model experiments have been performed^{8,41} in an attempt to estimate this strength, and it is generally agreed that the thermohaline circulation will act in such a way as to inhibit descent into the hard snowball state, thermohaline circulation behaviour during the Neoproterozoic must be considered ill-constrained at present. This requires that a range of values for the parameter F_{21} be considered. Additional inhibition to snowball Earth formation in the actual climate system may derive from the action of the wind-driven circulation¹⁵ and the detailed characteristics related to sea-ice formation in the tropics^{14,15} where incoming solar radiation is especially intense. Such additional processes have been fully incorporated in a recently published simulation⁸ of Neoproterozoic climate that has been performed using the NCAR CSM 1.4 model.

The 'two box model' of the carbon cycle that is assumed to respond to climate variability in the way described by the above relationships is coupled to a global energy balance model (EBM) that is itself coupled to a thermodynamic sea-ice module and a detailed model of continental-scale glaciation. The global EBM is based upon that of North *et al.*⁴² as modified by Deblonde *et al.*⁴³ to include a sea-ice component. Because carbon dioxide is a well-mixed trace gas, when an increase of the remineralization flux occurs as a consequence of decreasing temperature and the resulting increase in carbon dioxide dissolved in the ocean

is partly partitioned into the atmosphere, there is no need to consider atmospheric variations in the spatial distribution of this greenhouse gas.

The model of continental-scale glaciation is based upon the use of the standard Glen flow law for the rheology and the shallow-ice approximation for the dynamics. Because the topography of the Neoproterozoic continents is unknown, the ice cover on the individual continental fragments is assumed to develop on continental masses with a fixed initial freeboard of 400 m. As ice flows from the land to the sea under the action of the gravitational force, calving is assumed to act so as to eliminate ice beyond the continental termini. Processes such as fast flow which are strongly conditioned by basal hydrology and sediment deformation are eliminated from the model we employed. The isostatic adjustment process that acts to modify the surface elevation of the ice sheet and thus affects the surface mass balance is incorporated by assuming the process to act in a simple damped return-to-equilibrium fashion with an assumed relaxation time of 4,000 years, in acceptable accord with that appropriate for the Earth's present-day radial viscoelastic structure⁴⁴. This is expected to be an excellent approximation because, although the rheology of the planetary interior is strongly temperature-dependent, the temperature of the interior has not changed significantly since Neoproterozoic time⁴⁵.

The climate forcing that is assumed to drive the advances and retreats of both continental ice sheets and sea-ice cover is obtained from the surface temperature field delivered by the EBM, assuming an annually averaged precipitation rate of $\sim 0.7 \text{ m yr}^{-1}$, which is appropriately diminished owing to the action of the elevation desert effect, which acts so as to reduce the precipitation rate in regions of high topographic relief such as for continents covered by substantial thicknesses of glacial ice. These assumptions are the same as those previously employed in the application of this model to the climate of the Neoproterozoic⁷. The physical climate component of the carbon-cycle-coupled model has been thoroughly exercised in a number of recent analyses of climate variations through the Phanerozoic eon^{46,47} and in spite of its modest degree of complexity, compared well against expectations based upon the GEOCARB model⁴⁸ of past variations in atmospheric carbon dioxide.

40. Garcia, H. E. & Gordon, L. I. Oxygen solubility in sea water: Better fitting equations. *Limnol. Oceanogr.* **37**, 1307–1312 (1992).
41. Poulsen, C. J., Pierrehumbert, R. T. & Jacob, R. L. Impact of ocean dynamics on the simulation of the Neoproterozoic "snowball Earth". *Geophys. Res. Lett.* **28**, doi:10.1029/2000GL012058 (2001).
42. North, G. R., Mengel, J. G. & Short, D. A. Simple energy balance model resolving the seasons and continents: application to the astronomical theory of the ice ages. *J. Geophys. Res.* **88**, 6576–6586 (1983).
43. Deblonde, G., Peltier, W. R. & Hyde, W. T. Simulations of continental ice sheet growth over the last glacial-interglacial cycle: experiments with a one level seasonal energy balance model including seasonal ice albedo feedback. *Glob. Planet. Change* **98**, 37–55 (1992).
44. Peltier, W. R. Postglacial variations in the level of the sea: implications for climate dynamics and solid-earth geophysics. *Rev. Geophys.* **36**, 603–689 (1998).
45. Butler, S. L., Peltier, W. R. & Costin, S. O. Numerical models of the earth's thermal history: Effects of inner core solidification and core potassium. *Phys. Earth Planet. Inter.* **152**, 22–42 (2005).
46. Hyde, W. T., Crowley, T. J., Tarasov, L. & Peltier, W. R. The Pangean ice-age: Studies with a coupled climate-ice sheet model. *Clim. Dyn.* **15**, 619–629 (1999).
47. Hyde, W. T., Grossman, E. L., Crowley, T. J., Pollard, D. & Scotese, C. R. Siberian glaciation as a constraint on Permian–Carboniferous CO₂ levels. *Geology* **34**, 421–424 (2006).
48. Berner, R. A. GEOCARBSULF: A combined model for Phanerozoic atmospheric O₂ and CO₂. *Geochim. Cosmochim. Acta* **70**, 5653–5664 (2006).

Engraftment of connexin 43-expressing cells prevents post-infarct arrhythmia

Wilhelm Roell^{1,2*}, Thorsten Lewalter^{3*}, Philipp Sasse^{1*}, Yvonne N. Tallini⁶, Bum-Rak Choi⁷, Martin Breitbach¹, Robert Doran⁶, Ulrich M. Becher^{1,3}, Seong-Min Hwang⁸, Toktam Bostani^{1,2}, Julia von Maltzahn⁴, Andreas Hofmann⁵, Shaun Reining⁶, Britta Eiberger⁴, Bethann Gabris⁸, Alexander Pfeifer⁵, Armin Welz², Klaus Willecke⁴, Guy Salama⁸, Jan W. Schrickel^{1,3}, Michael I. Kotlikoff⁶ & Bernd K. Fleischmann¹

Ventricular tachyarrhythmias are the main cause of sudden death in patients after myocardial infarction. Here we show that transplantation of embryonic cardiomyocytes (eCMs) in myocardial infarcts protects against the induction of ventricular tachycardia (VT) in mice. Engraftment of eCMs, but not skeletal myoblasts (SMs), bone marrow cells or cardiac myofibroblasts, markedly decreased the incidence of VT induced by *in vivo* pacing. eCM engraftment results in improved electrical coupling between the surrounding myocardium and the infarct region, and Ca^{2+} signals from engrafted eCMs expressing a genetically encoded Ca^{2+} indicator could be entrained during sinoatrial cardiac activation *in vivo*. eCM grafts also increased conduction velocity and decreased the incidence of conduction block within the infarct. VT protection is critically dependent on expression of the gap-junction protein connexin 43 (Cx43; also known as Gja1): SMs genetically engineered to express Cx43 conferred a similar protection to that of eCMs against induced VT. Thus, engraftment of Cx43-expressing myocytes has the potential to reduce life-threatening post-infarct arrhythmias through the augmentation of intercellular coupling, suggesting autologous strategies for cardiac cell-based therapy.

Cell transplantation has emerged as a potential treatment strategy for heart failure secondary to acute or chronic ischaemic heart disease (reviewed in ref. 1). Currently two autologous cell types, namely bone marrow (BM) cells and SMs, are used in clinical trials in patients after myocardial infarction (reviewed in ref. 2). Both of these cell types seem to provide only modest improvement of contractile heart function^{3–5} because neither BM cells^{6,7} nor SMs⁸ adopt a cardiac cell fate or couple electrically with the host myocardium^{6,9}. Moreover, VT has been reported in several of the patients transplanted with SMs^{5,10}, raising concerns as to whether engraftment of skeletal-muscle-derived cells enhances the risk of VT, the most frequent cause of sudden death after myocardial infarction^{11,12}. This led to the recommendation that cardiac SM engraftment be limited to patients with an implantable cardioverter defibrillator¹³. Transplantation of eCMs in animal models results in effective engraftment with expression of gap-junction proteins and modest augmentation of heart performance^{14,15}. Here we show that eCMs confer marked protection against ventricular arrhythmias by enhancing intercellular electrical coupling within the engrafted infarct. We also show that genetic modification of SMs to express Cx43 not only eliminates the pro-arrhythmogenic effect of SM transplantation but also provides arrhythmia protection that is equivalent to that of engraftment with eCMs. Thus, augmentation of intercellular coupling by cell-based therapy may be an effective therapeutic strategy for the prevention of post-infarct VT.

Vulnerability testing *in vivo*

We assessed electrical vulnerability *in vivo* by burst (Fig. 1a and Supplementary Fig. 1b) and extrastimulus pacing (Supplementary

Fig. 1c) protocols in mice with left ventricular infarcts (Fig. 1b, left panel) 11–14 days after injury and transplantation of SMs, BM cells, cardiac myofibroblasts or eCMs. *In vivo* pacing protocols induced VT in 38.9% ($n = 18$) of non-infarcted control mice (HIM:OF1 and CD1 background) (Figs 1a and 2h, and Supplementary Fig. 1b). In contrast, monomorphic or polymorphic, self-terminating VT (Fig. 1c and Supplementary Fig. 1c) could be evoked in 96.4% of infarcted and vehicle-injected (sham-injected) mice ($n = 28$; Fig. 2h), although only minor alterations to the electrocardiogram (ECG) were observed under baseline conditions (Fig. 1d, left panel, and Supplementary Fig. 1a). VT was proved by the presence of atrio-ventricular dissociation at the His bundle level (Fig. 1c, lower trace). Thus, localized myocardial infarction markedly enhances VT inducibility in mice, allowing a systematic evaluation of the consequences of the engraftment of different cell types on electrical vulnerability.

Engraftment of SMs enhances the severity of VT

We next tested the effect of SM transplantation on electrical stability of infarcted hearts by injecting embryonic enhanced green fluorescent protein (EGFP)-positive (ref. 16) skeletal-muscle-derived cells (Supplementary Fig. 1d). These cells formed a confined layer within the infarct two weeks after injury (Fig. 1b, right panel, and Fig. 1f, g). The engrafted SMs were elongated and multinucleate with central nuclei (Fig. 1g, inset) and expressed the skeletal muscle marker nebulin (Fig. 1h), indicating ongoing skeletal differentiation and, as reported earlier¹⁷, no evidence of cardiac ‘transdifferentiation’. SM-engrafted hearts did not express the gap-junction protein Cx43 in the graft region. Similarly to untreated infarcted animals, ST- and T-wave alterations were minor, but a split QRS complex without

¹Institute of Physiology I, Life and Brain Center, ²Department of Cardiac Surgery, ³Department of Internal Medicine II, ⁴Institute of Genetics, ⁵Institute of Pharmacology, University of Bonn, Bonn 53105, Germany. ⁶Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853-6401, USA. ⁷Cardiovascular Research Center, Rhode Island Hospital and Brown Medical School, Providence, Rhode Island 02903, USA. ⁸Department of Cell Biology and Physiology, University of Pittsburgh, School of Medicine, Pittsburgh, Pennsylvania 15261, USA.

*These authors contributed equally to this work.

significant QRS prolongation was commonly observed (Fig. 1d, right panel, and Supplementary Table 1), suggesting inhomogeneity of ventricular activation. In all except one mouse containing EGFP-positive SMs ($n = 16$), VT could be evoked by pacing protocols (Figs 1e and 2h), and in 25% of mice these were found to degenerate into polymorphic VT and ventricular fibrillation (sustained arrhythmias; Fig. 1e). Although the induction of VT was almost 100% in both sham-injected and SM-engrafted mice the incidence of sustained arrhythmias was significantly increased in SM-transplanted mice ($P = 0.0134$). The equivalent incidence of arrhythmias in non-engrafted or SM-engrafted mice indicates that transplantation of non-electrically coupled cells does not decrease a major risk factor associated with myocardial infarction.

eCM engraftment decreases vulnerability to VT

Transplanted eCMs engrafted in a more diffuse pattern (Fig. 2b, e) than SMs, resulting in an apparently less thickened myocardial wall (compare right panel of Fig. 1b with Fig. 2a, b). The engrafted EGFP-positive SMs and eCMs could be clearly distinguished from the native myocardium, as confirmed by anti-EGFP staining (Supplementary Fig. 1e–h). Quantitative morphometry indicated engraftment in the range 3,000–20,000 eCMs ($n = 11$, median 4,480), similar to results reported earlier¹⁵. Moreover, immunostaining revealed Cx43 expression between the engrafted, differentiated eCMs (Fig. 2g). Occasionally, contact between graft and host myocardium and

Cx43 staining between transplanted eCMs and native cardiomyocytes in the border zone of the infarct was observed (Fig. 2f, g). Despite the relatively modest degree of engraftment, transplantation of eCMs, which increases the survival rate of infarcted mice¹⁵, markedly improved electrical stability. In contrast to the sham-injected group, eCM engraftment caused a marked decrease in the incidence of VT that was similar to values of the non-infarcted control group (35.7%, $n = 22$, and 38.9%, $n = 18$, for eCMs and control, respectively; Fig. 2d, h). Quantitative morphometry did not reveal an association between contact of the grafted eCMs with the host myocardium in the border zone of the infarct and VT protection, because, in a series of combined electrophysiological investigation and morphometry, in two-thirds of protected mice ($n = 6$) no contact was found; VT protection was also observed in hearts with a relatively low degree of engraftment ($n = 3$). Surface ECGs were similar to those of controls; in contrast to the SM group, no deformation of the QRS complex was seen (Fig. 2c and Supplementary Table 1). The observed difference in VT incidence is unlikely to have been due to engrafted non-muscle cells, because both SM (Supplementary Fig. 1d) and eCM¹⁵ preparations contain similar percentages of fibroblasts. Despite the markedly different consequences of SM and eCM transplantation with respect to arrhythmogenicity, both procedures augmented left ventricular function to a similar extent (Fig. 2i), excluding heart failure as the mechanism underlying differences in electrical stability. We also tested the effect of transplanting BM cells,

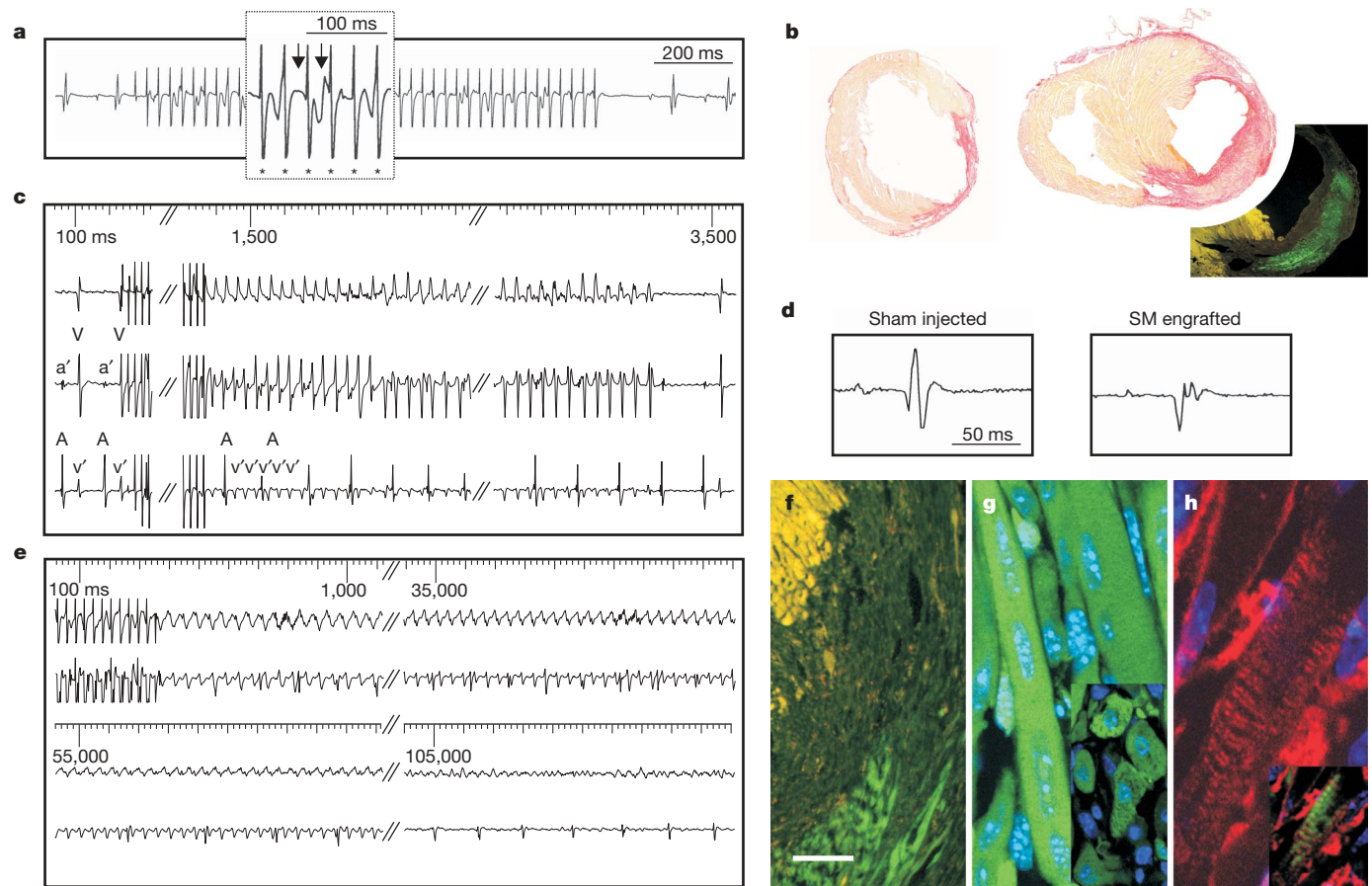


Figure 1 | VT in control, sham-injected and SM-engrafted hearts. **a**, Surface ECG from a control mouse including magnified inset of the burst stimulation (asterisks) protocol. Arrows indicate ventricular response; VT is not evoked. **b**, Sirius red staining proves transmurular fibrotic scar 2 weeks after sham injection (left) or engraftment of SMs into the scar (right; inset shows EGFP-positive SMs). **c**, Burst stimulation induces self-terminating VT with atrio-ventricular dissociation in sham-injected mouse. Top trace, surface ECG; middle trace, ventricular apex electrogram; bottom trace, His-level electrogram. A, atrium; V, ventricle; a', atrial farfield; v', ventricular

farfield. **d**, ECG recording from sham-injected mouse and SM-engrafted mouse. **e**, Burst pacing induces sustained VT in SM-engrafted mouse; note constant atrio-ventricular dissociation. Upper trace, surface ECG; lower trace, His-level electrogram. **f**, Engrafted EGFP-positive (green) SMs; native myocardium is marked by yellow autofluorescence. **g**, Engrafted SM fibres are elongated and multinucleate with central nuclei (inset). **h**, Transplanted EGFP-positive SMs (green in inset); double cross-striation shown with nebulin staining (red). Scale bar, 1.3 mm (**b**), 150 μ m (**f**), 13 μ m (**g**), 16 μ m (**g**, inset), 7 μ m (**h**), 18 μ m (**h**, inset).

a second cell type currently used in clinical trials, on electrical stability. Engraftment of BM cells ($n = 17$) did not decrease the incidence or severity of induced VT (Fig. 2h). Similar results were also obtained after transplantation of cardiac myofibroblasts ($n = 9$; Fig. 2h and Supplementary Fig. 3a–c).

eCM grafts couple to native myocardium *in vivo*

We proposed that augmented electrical stability after eCM transplantation could result from the electrical integration of transplanted

cardiomyocytes, because eCMs have previously been shown to couple to native myocytes after transplantation into the non-injured myocardium *ex vivo*¹⁸. We therefore examined whether eCMs integrate functionally within infarcted tissues *in vivo* by transplanting cells isolated from mouse hearts expressing the genetically encoded fluorescent Ca^{2+} indicator GCaMP2 (ref. 19) (Fig. 3f) or transduced with lentivirus expressing GCaMP2 (Fig. 3a–e and Supplementary Fig. 1i, j). Both approaches allowed us to track the function of implanted cells *in vivo*, including the degree of activation by the surrounding myocardium during normal sinoatrial rhythm, as well as the amplitude and time course of Ca^{2+} transients in transplanted cells. *In vivo* optical recordings of Ca^{2+} -dependent fluorescence in open-chested mice 9–16 days after transplantation of eCMs indicated that Ca^{2+} signals from engrafted cells were entrained to the heart rhythm in infarct areas. Ca^{2+} signals were often coupled below the sinus rate (1:2 or 1:4) and reflected a conduction delay as indicated by simultaneous ECG and Ca^{2+} -dependent fluorescence recordings (Fig. 3c and Supplementary Video 1). In three such experiments, the delay between the QRS minimum and activation of the GCaMP2⁺ cells was 70.1 ± 9.5 ms (mean \pm s.e.m.); this value is an overestimate because of the delay between electrical excitation and Ca^{2+} and the on-rate of GCaMP2 (ref. 19). Engrafted eCMs could be stimulated by pacing from an electrode placed outside the infarct

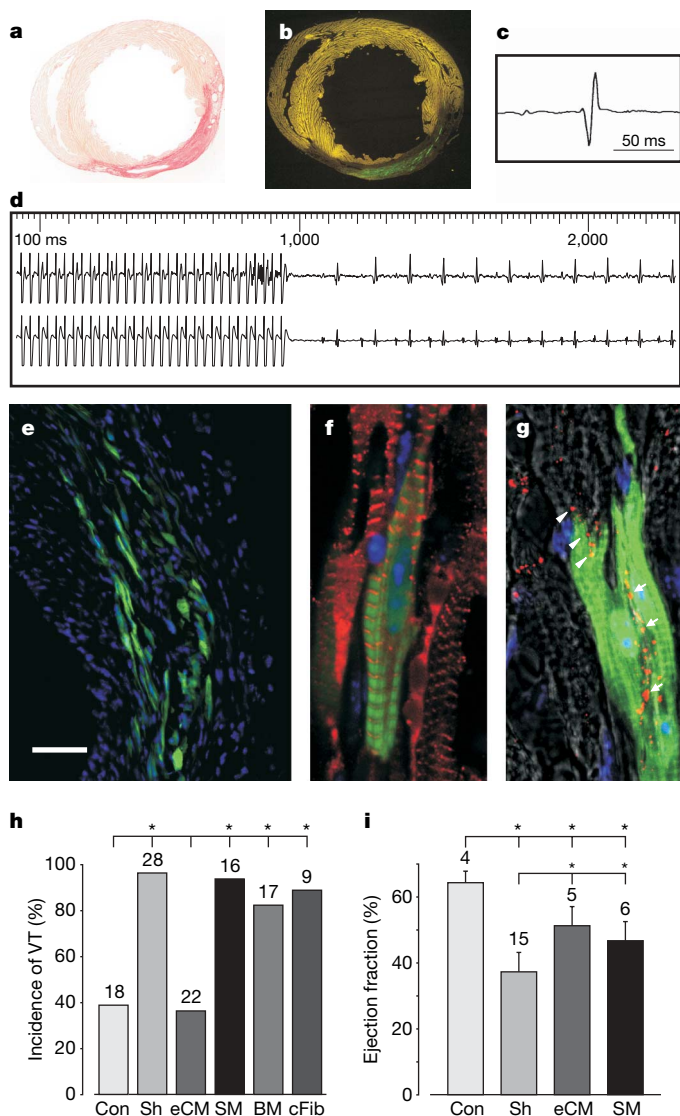


Figure 2 | VT protection in eCM-engrafted hearts. **a**, Sirius red staining. **b**, EGFP-positive eCMs engraft in the infarct and extend to the border zone. **c**, ECG recording of eCM-engrafted mouse. **d**, Ventricular burst stimulation does not induce VT. Upper trace, surface ECG; lower trace, His-level electrogram. **e**, EGFP-positive eCMs integrate into the infarct. **f**, EGFP-positive eCMs are striated (α -actinin staining, red) and are in direct contact with EGFP-negative host cardiomyocytes. **g**, Cx43 staining (red) illustrates gap-junction formation between engrafted EGFP-positive eCMs (arrows) and with native cardiomyocytes (arrowheads). **h**, Summary of VT inducibility. Note strongly elevated susceptibility to VT induction in sham-injected group (Sh) compared with control group (Con, $P < 0.0001$); eCM engraftment reduces VT inducibility compared with sham injection ($P < 0.0001$). Conversely, VT remain frequent after transplantation of SMs, BM and cardiac myofibroblasts (cFib) ($P = 0.0011$ and 0.0005 ; 0.015 and 0.008 ; 0.019 and 0.016 versus control and eCMs, respectively). **i**, Improvement of left ventricular ejection fraction after eCM or SM transplantation. Asterisk, $P < 0.005$. Numbers above bars indicate n ; error bars show s.d. Scale bar, 1.6 mm (**a**, **b**), 60 μm (**e**), 5 μm (**f**), 11 μm (**g**).

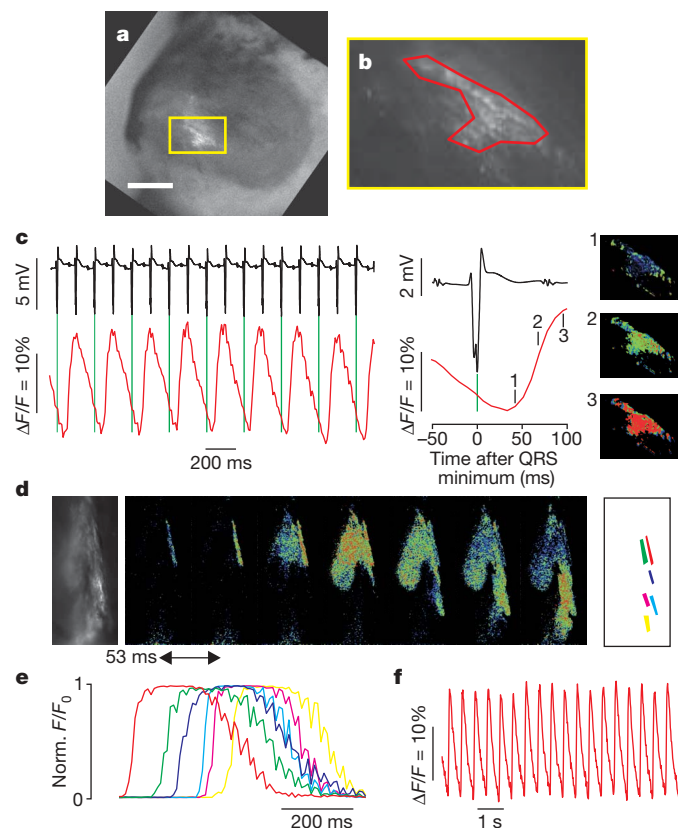


Figure 3 | Entrainment of engrafted eCMs *in vivo*. **a**, Image of a heart (taken *ex vivo*) with engrafted GCaMP2⁺ cells in the infarct (dark area). **b**, Magnified view of the yellow box in **a**. **c**, Combined *in vivo* recordings (left) of ECG (black) and Ca^{2+} transients (red, area marked in **b**) reveal 2:1 coupling of native myocardium with engrafted cells. Averaging of coupled QRS (green lines) and Ca^{2+} signals yields their temporal correlation (middle). Pseudocolour images of GCaMP2 fluorescence at time points indicated by numbers (right). **d**, **e**, High-resolution Ca^{2+} fluorescence from engrafted GCaMP2⁺ cells (**d**, left) in a Langendorff-perfused heart. Image series (**d**, middle) demonstrates coupling of engrafted cells, with sequential activation (**e**) of individual regions (**d**, right). **f**, Entrainment of Ca^{2+} fluorescence *in vivo* to 2-Hz stimulation from electrode in normal myocardium (intrinsic heart rate slowed with the use of deep anaesthesia). Scale bar, 1.5 mm (**a**), 360 μm (**b**, **d**), 3.2 mm (**c**).

area, resulting in synchronized Ca^{2+} transients at the stimulus rate within the infarct zone (Fig. 3f). In addition, transplanted cells showed uncoupled activity that conducted slowly throughout the engrafted myocytes, although ectopic beats or arrhythmias associated with this activity were never observed. In Langendorff-perfused hearts, in which heart motion could be constrained, high-resolution imaging of engrafted GCaMP2^+ cells in the infarct revealed their sequential activation and hence electrical coupling (Fig. 3d, e). Thus, activation signals were transferred across the border zone to cells within the infarct region, over a distance of as much as 1 mm, and transplantation of eCMs resulted in two types of functional coupling *in vivo*: heterologous coupling between transplanted cells and host cardiomyocytes, and autologous or regional coupling within islands of transplanted cells. Moreover, in Langendorff-perfused hearts loaded with Rhod-2, Ca^{2+} signals crossed the infarct border and were recorded within the infarct zone (left panels of Supplementary Fig. 2c, d, and lower panels of Supplementary Fig. 2e) in contrast to sham-injected hearts in which the Ca^{2+} wave largely bypassed the infarct, with minimal Ca^{2+} transients within the infarct zone (right panels of Supplementary Fig. 2c, d, and upper panels of Supplementary Fig. 2e). We reasoned that entrained Ca^{2+} signals in transplanted eCMs and a decreased incidence of conduction block of Rhod-2 Ca^{2+} transients reflect enhanced intercellular electrical coupling that probably underlies the marked decrease in arrhythmia vulnerability after eCM transplantation.

Enhanced conduction in eCM-engrafted infarcts

Although the mechanisms underlying post-infarction re-entry arrhythmias are not completely understood, wave breaks at the boundary of an anatomic block^{20,21} and focal activity including early and delayed afterdepolarizations arising from the infarct border zone²² are the characteristic electrophysiological features of arrhythmogenesis after infarction. We performed optical mapping by using Langendorff perfusion²³ in sham-injected and eCM-engrafted hearts two weeks after infarct or engraftment and observed both wave breaks and focal activity (Fig. 4b, upper panels). Under sinus rhythm and unipolar stimulation from outside the infarct zone, action potentials propagated transmurally, emerged on the epicardium, and travelled around the infarct zone in sham-injected hearts ($n = 5$; Fig. 4a, upper panels, and Supplementary Fig. 2a, upper panels); pacing revealed prominent conduction blocks at the boundaries of infarcted regions (Fig. 4b, upper left panel). In contrast, action potentials propagated into the infarct zone in eCM-engrafted hearts (Fig. 4b, lower panel) and were recorded within areas of the infarct ($n = 5$, Fig. 4a, lower panels, and Supplementary Fig. 2a, upper panels). Additionally, non-engrafted hearts showed a higher incidence of focal activity emanating from the infarct border zone (Fig. 4b, upper right panel); ectopic beats were observed in 38% of sham-injected ($n = 8$) and 14% of eCM-engrafted ($n = 7$) hearts. In addition, premature ventricular stimuli designed to evoke re-entry phenomena induced prominent spiral electrical waves in sham-injected hearts ($n = 3$; Fig. 4d and Supplementary Video 2), whereas the same protocols did not induce re-entry waves in eCM-engrafted hearts ($n = 5$, Supplementary Video 3), indicating that the conditions required to elicit stable re-entrant circuit movement²⁴ were eliminated in these hearts. Thus, eCM engraftment decreases the incidence of conduction block and wave breaks, and suppresses border-zone focal activity, the two hallmarks of post-infarct ventricular arrhythmia.

To determine whether these effects result from augmented intercellular electrical coupling, we measured local conduction velocities from the activation time points of optical action potentials²⁵ and found that conduction velocity within the infarct was increased more than fourfold in eCM-engrafted hearts (Fig. 4c), which is consistent with a marked increase in intercellular coupling. Further evidence of the critical nature of gap-junctional conductance within the infarct zone was obtained by perfusion of hearts with the connexin blocker

carbenoxolone (0.5 mM). After perfusion of the drug, the conduction velocity decreased more rapidly and to a greater extent in the infarct zone than in the intact myocardium of sham-injected hearts ($29 \pm 9\%$ versus $53 \pm 4\%$, $n = 4$), indicating a contribution by gap junctions to conduction, as well as a lower conduction reserve, within the infarct zone. Anti-Cx43 immunostaining and western blotting revealed significantly lower Cx43 expression in cardiac

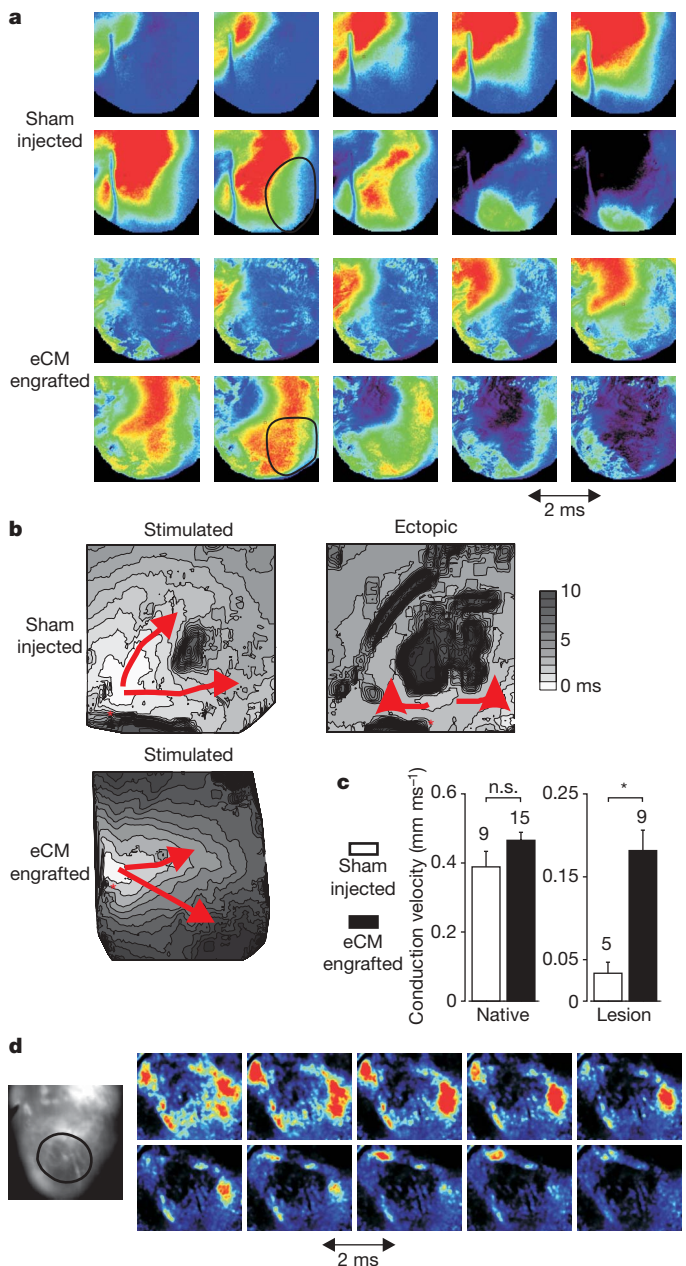


Figure 4 | eCM engraftment improves conduction and prevents re-entry in Langendorff-perfused hearts. **a**, Sequential images from a series showing the first derivative (2-ms interval) of action-potential propagation during pacing. The infarct region is outlined by a black circle. **b**, Representative isochronal maps depicting the activation wavefront and conduction delays near the infarct border zone. The activation bypasses the sham-injected infarct (upper left), or initiates from an ectopic focus at the border zone (right), whereas it proceeds through the eCM-injected lesion (lower left). Scale bar indicates local activation times. **c**, Local conduction velocities. There is a more than fourfold increase within eCM-engrafted infarcts ($P < 0.001$; numbers above bars indicate n ; error bars show s.e.m.). **d**, Sequential di-4-ANNEPS fluorescence images show induction of spiral waves in a sham-injected heart (images cropped around infarct zone). The infarct region is ringed in the leftmost panel.

myofibroblasts than in eCMs and adult heart (Supplementary Fig. 3d–f). Although cardiac myofibroblasts do not themselves provide sufficient VT protection (Fig. 2h), electrical coupling of engrafted eCMs that do not extend to the border zone probably reflects the contribution of cardiac myofibroblasts to the electrotonic transfer of action potentials across the border zone²⁶ and improved electrical coupling within the infarct zone. It is possible, however, that arrhythmia protection results from tonic effects on border-zone myocytes (for example, changes in resting membrane potential) associated with the engraftment of electrically coupled eCMs. To examine this possibility, we determined conduction velocities and action-potential rise times, parameters that would probably be altered in the event of dominant effects of border-zone cardiomyocytes, in the border zone of infarcts from optical mapping data. Conduction velocities were very similar in sham-injected (0.178 ± 0.011 mm ms⁻¹, $n = 3$) and in eCM-engrafted (0.164 ± 0.037 mm ms⁻¹, $n = 4$) hearts. In addition, action-potential rise times were almost identical in border-zone myocytes of sham-injected (5.6 ± 1.5 ms, $n = 3$) and eCM-engrafted (5.7 ± 1.7 ms, $n = 4$) hearts. However, these measurements do not exclude alterations in subregions of the border

zone that could give rise to the decreased focal activity that we observed. Taken together, these data indicate that eCM engraftment markedly enhances intercellular electrical conduction within the infarct zone, resulting in the elimination or reduction of electrical events that initiate re-entry arrhythmias (focal activity and conduction blocks) and provide a mechanistic basis for the markedly decreased electrical vulnerability produced by eCM engraftment.

SM-Cx43 engraftment confers arrhythmia protection

We directly assessed the role of enhanced electrical coupling through Cx43 in arrhythmia protection, by engrafting SMs from transgenic mice expressing Cx43 (SM-Cx43⁺) under the control of a skeletal muscle promoter (see Supplementary Methods and Supplementary Fig. 4a–e). This was also intended to exclude the possibility that engraftment of eCMs indirectly enhances conduction by paracrine mechanisms²⁷ and to explore the feasibility of generating an accessible and autologous cell source for potential clinical applications. Functional gap junctions are able to form between SM-Cx43⁺ myotubes, as shown by dye-transfer studies *in vitro* (Supplementary Fig. 4f). SM-Cx43⁺ stably engrafted into the infarcted myocardium (Fig. 5a, b) differentiated, forming multinucleate cells with a distinct cross-striation (Fig. 5f), and expressed Cx43 (Fig. 5f, g, and Supplementary Fig. 4g). Surface ECGs recorded from these mice showed only minor alterations compared with ECGs of control and sham-injected animals (Fig. 5c and Supplementary Table 1), and QRS splitting was observed after transplantation of wild-type SMs (Fig. 1d) but not SM-Cx43⁺ (Fig. 5c). Stable engraftment of SM-Cx43⁺ markedly improved electrical stability (Fig. 5d, h), because VT was induced in only 37.5% ($n = 16$) of SM-Cx43⁺-engrafted mice, an improvement in electrical stability almost identical to that achieved by the transplantation of eCMs. In contrast, 100% of mice transplanted with SMs not expressing Cx43 from control littermates (SM-Cx43⁻) developed VT ($n = 13$; Fig. 5e, h). These experiments clearly demonstrate that Cx43 expression is necessary for protection from arrhythmia and exclude paracrine mechanisms.

Discussion

Taken together, our results show that cardiomyocyte transplantation has the potential to impart electrical stability to the injured heart, thereby markedly reducing the major factor leading to sudden death. This protective effect is independent of the documented modest augmentation of left ventricular function and is associated with improved electrical coupling within the infarct by the engraftment of Cx43-expressing eCMs. This enhanced coupling reduces vulnerability to VT by decreasing the incidence of conduction block within the infarct and/or by a modulatory effect on border-zone cardiomyocytes. Although conduction velocity and action-potential rise time are not altered in the border zone, subtle effects cannot be excluded and are supported by the observation of decreased ectopic activity after eCM engraftment. Expression of the cardiac gap-junction protein Cx43 is the critical factor underlying augmented intercellular electrical conduction and protection from arrhythmia, because engraftment of Cx43-expressing SMs protects against VT induction. Because autologous SM muscle can be isolated from humans and protection is achieved with relatively small numbers of engrafting cells, transplantation of SMs in combination with Cx43 gene transfer represents a promising therapeutic strategy to decrease the risk of potentially fatal arrhythmias.

METHODS SUMMARY

Harvesting of cells and transplantation. Single cells were dissociated from embryonic hearts (embryonic development day (E)14.5–E16.5) and skeletal muscles (E18.5/E19.5) of human cardiac α -actin-EGFP transgenic mice¹⁶ (CD1 and HIM:OF1 background). Cx43⁺ SMs and their controls were obtained from MCK-tTA/Cx43TetOEGFP embryos and their negative littermates (C57/Bl6 background). Cardiac myofibroblasts were obtained by pre-plating of embryonic heart cells (E14.5/E15.5). BM cells were aspirated from hindlimbs of adult wild-type CD1 mice²⁸. Dissociated cells (10^5 SMs, eCMs or cardiac

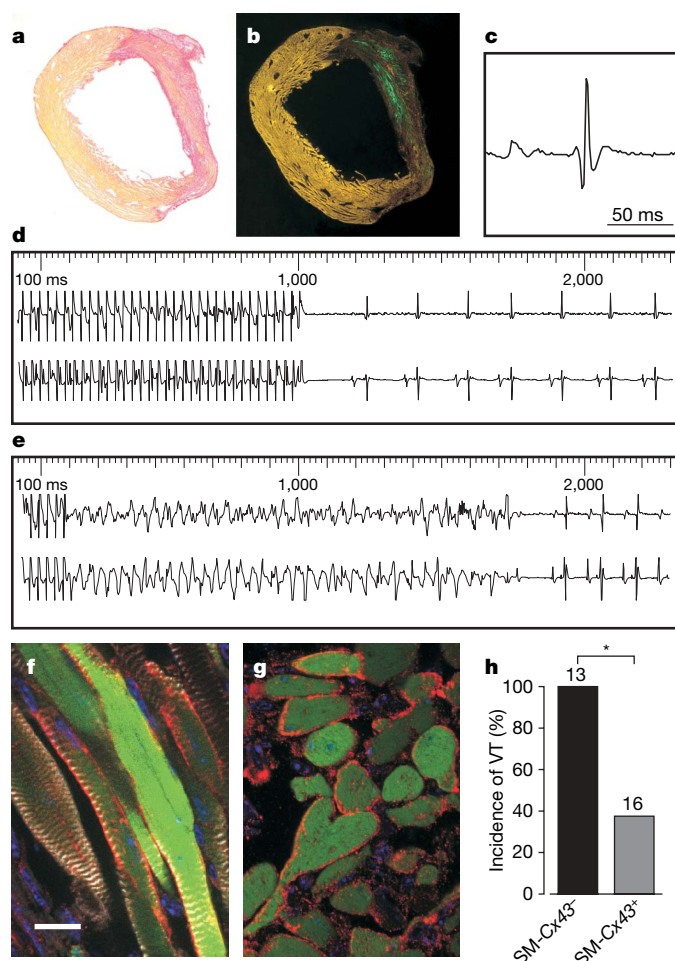


Figure 5 | VT protection in Cx43-expressing SM-engrafted hearts. a–c, Sirius red staining (a), native EGFP fluorescence (green, b) and ECG recording (c) in SM-Cx43⁺-engrafted hearts. d, e, Ventricular burst stimulation induces no VT in SM-Cx43⁺-engrafted mouse (d), but transient VT in SM-Cx43⁻-engrafted mouse (e). Upper traces, surface ECG; lower traces, His-level electrogram. f, g, EGFP-positive (green) SM-Cx43⁺ engraft stably into the infarcted heart. Myotubes are multinucleate and striated (f, α -actinin staining in white). Cx43 immunostaining (red) shows punctate expression at the cell surface in longitudinal section (f) and cross-section (g). h, VT inducibility of SM-Cx43⁺-engrafted and SM-Cx43⁻-engrafted hearts ($P < 0.0004$). Numbers above bars indicate n . Scale bar, 1.2 mm (a, b), 15 μ m (f, g).

myofibroblasts, or 3×10^6 BM cells) or vehicle (sham-injected) were injected into the cryolesioned anterior-lateral left ventricular wall of recipient mice²⁹. Non-operated mice were used as a control group.

In vivo electrophysiology. ECG parameters and inducibility of VT were evaluated by a blinded investigator 11–14 days postoperatively by transvenous electrophysiological investigation by using burst and premature-beat stimulation³⁰. VT was defined as at least four ectopic ventricular beats characterized by atrio-ventricular dissociation.

Histology and immunohistochemistry. After electrophysiological investigation, hearts were imaged by fluorescence macroscopy, fixed in paraformaldehyde and cryosectioned. The extent of myocardial lesions was determined by staining with Sirius red. Engrafted cells were identified by their native EGFP fluorescence and characterized by immunohistochemistry.

In vivo imaging and ex vivo optical mapping. *GCaMP2⁺* eCMs were derived from transgenic embryos¹⁹ or obtained by lentiviral transduction of plated eCMs (CMV-GCaMP2)³¹. *In vivo* experiments were performed in ventilated and thoracotomized mice. *GCaMP2* signals from transplanted cells were recorded with a fluorescence microscope (OV100 or MVX10; Olympus) and an electron-multiplied, charge-coupled-device camera at frame rates of 67–128 Hz.

For *ex vivo* recordings, hearts were Langendorff perfused, suspended in an optical chamber and labelled with the membrane-potential-sensitive dye di-4-ANEPPS or the fluorescent calcium indicator Rhod-2. Voltage or Ca^{2+} signals were recorded with a complementary-metal-oxide semiconductor (CMOS) camera at 1,000 Hz. Conduction velocity within the infarct or normal regions was measured under point stimulation at a cycle length of 200 ms for 20 beats.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 June; accepted 28 September 2007.

- Laflamme, M. A. & Murry, C. E. Regenerating the heart. *Nature Biotechnol.* **23**, 845–856 (2005).
- Murry, C. E., Field, L. J. & Menasche, P. Cell-based cardiac repair: reflections at the 10-year point. *Circulation* **112**, 3174–3183 (2005).
- Wollert, K. C. *et al.* Intracoronary autologous bone-marrow cell transfer after myocardial infarction: the BOOST randomised controlled clinical trial. *Lancet* **364**, 141–148 (2004).
- Janssens, S. *et al.* Autologous bone marrow-derived stem-cell transfer in patients with ST-segment elevation myocardial infarction: double-blind, randomised controlled trial. *Lancet* **367**, 113–121 (2006).
- Smits, P. C. *et al.* Catheter-based intramyocardial injection of autologous skeletal myoblasts as a primary treatment of ischemic heart failure: clinical experience with six-month follow-up. *J. Am. Coll. Cardiol.* **42**, 2063–2069 (2003).
- Murry, C. E. *et al.* Haematopoietic stem cells do not transdifferentiate into cardiac myocytes in myocardial infarcts. *Nature* **428**, 664–668 (2004).
- Nygren, J. M. *et al.* Bone marrow-derived hematopoietic cells generate cardiomyocytes at a low frequency through cell fusion, but not transdifferentiation. *Nature Med.* **10**, 494–501 (2004).
- Reinecke, H., Poppa, V. & Murry, C. E. Skeletal muscle stem cells do not transdifferentiate into cardiomyocytes after cardiac grafting. *J. Mol. Cell. Cardiol.* **34**, 241–249 (2002).
- Leobon, B. *et al.* Myoblasts transplanted into rat infarcted myocardium are functionally isolated from their host. *Proc. Natl Acad. Sci. USA* **100**, 7808–7811 (2003).
- Hagege, A. A. *et al.* Skeletal myoblast transplantation in ischemic heart failure: long-term follow-up of the first phase I cohort of patients. *Circulation* **114**, 1108–1113 (2006).
- Solomon, S. D. *et al.* Sudden death in patients with myocardial infarction and left ventricular dysfunction, heart failure, or both. *N. Engl. J. Med.* **352**, 2581–2588 (2005).
- Henkel, D. M. *et al.* Ventricular arrhythmias after acute myocardial infarction: a 20-year community study. *Am. Heart J.* **151**, 806–812 (2006).
- Makkar, R. R., Lill, M. & Chen, P. S. Stem cell therapy for myocardial repair: is it arrhythmogenic? *J. Am. Coll. Cardiol.* **42**, 2070–2072 (2003).
- Soonpaa, M. H., Koh, G. Y., Klug, M. G. & Field, L. J. Formation of nascent intercalated disks between grafted fetal cardiomyocytes and host myocardium. *Science* **264**, 98–101 (1994).
- Roell, W. *et al.* Cellular cardiomyoplasty improves survival after myocardial injury. *Circulation* **105**, 2435–2441 (2002).
- Fleischmann, M. *et al.* Cardiac specific expression of the green fluorescent protein during early murine embryonic development. *FEBS Lett.* **440**, 370–376 (1998).
- Pouzet, B. *et al.* Is skeletal myoblast transplantation clinically relevant in the era of angiotensin-converting enzyme inhibitors? *Circulation* **104**, 1–223–1–228 (2001).
- Dowell, J. D., Rubart, M., Pasumarthi, K. B., Soonpaa, M. H. & Field, L. J. Myocyte and myogenic stem cell transplantation in the heart. *Cardiovasc. Res.* **58**, 336–350 (2003).
- Tallini, Y. N. *et al.* Imaging cellular signals in the heart *in vivo*: Cardiac expression of the high-signal Ca^{2+} indicator *GCaMP2*. *Proc. Natl Acad. Sci. USA* **103**, 4753–4758 (2006).
- Pertsov, A. M., Davidenko, J. M., Salomonsz, R., Baxter, W. T. & Jalife, J. Spiral waves of excitation underlie reentrant activity in isolated cardiac muscle. *Circ. Res.* **72**, 631–650 (1993).
- Takahashi, T. *et al.* Optical mapping of the functional reentrant circuit of ventricular tachycardia in acute myocardial infarction. *Heart Rhythm* **1**, 451–459 (2004).
- Grant, A. O., Hondeghem, L. M. & Katzung, B. G. Effects of droperidol on depolarization-induced automaticity, maximum upstroke velocity (v_{max}) and the kinetics of recovery of v_{max} in guinea-pig ventricular myocardium. *J. Pharmacol. Exp. Ther.* **205**, 193–203 (1978).
- Baker, L. C., London, B., Choi, B. R., Koren, G. & Salama, G. Enhanced dispersion of repolarization and refractoriness in transgenic mouse hearts promotes reentrant ventricular tachycardia. *Circ. Res.* **86**, 396–407 (2000).
- Allessie, M. A. *et al.* Experimental electrophysiology and arrhythmogenicity. Anisotropy and ventricular tachycardia. *Eur. Heart J.* **10** (Suppl. E), E2–E8 (1989).
- Salama, G., Kanai, A. & Efimov, I. R. Subthreshold stimulation of Purkinje fibers interrupts ventricular tachycardia in intact hearts. Experimental study with voltage-sensitive dyes and imaging techniques. *Circ. Res.* **74**, 604–619 (1994).
- Gaudesius, G., Miragoli, M., Thomas, S. P. & Rohr, S. Coupling of cardiac electrical activity over extended distances by fibroblasts of cardiac origin. *Circ. Res.* **93**, 421–428 (2003).
- Gnecchi, M. *et al.* Paracrine action accounts for marked protection of ischemic heart by Akt-modified mesenchymal stem cells. *Nature Med.* **11**, 367–368 (2005).
- Kolossov, E. *et al.* Engraftment of engineered ES cell-derived cardiomyocytes but not BM cells restores contractile function to the infarcted myocardium. *J. Exp. Med.* **203**, 2315–2327 (2006).
- Roell, W. *et al.* Cellular cardiomyoplasty in a transgenic mouse model. *Transplantation* **73**, 462–465 (2002).
- Kreuzberg, M. M. *et al.* Connexin30.2 containing gap junction channels decelerate impulse propagation through the atrioventricular node. *Proc. Natl Acad. Sci. USA* **103**, 5959–5964 (2006).
- Pfeifer, A., Ikawa, M., Dayn, Y. & Verma, I. M. Transgenesis by lentiviral vectors: lack of gene silencing in mammalian embryonic stem cells and preimplantation embryos. *Proc. Natl Acad. Sci. USA* **99**, 2140–2145 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Fuerst for providing the anti-nebulin antibody; L. Field for providing the transgenic α -MHC-EGFP mouse line; W. Bloch for advice on immunostaining; C. Schaffer and N. Nishimura for technical advice; R. Gilmour Jr and N. Otani for comments on the manuscript; C. Fuegmann for the preparation of cardiac myofibroblasts; and H. Begerau, M. Czechowski, B. Eixmann, F. Holst, H. Doerr, K. Granitz and C. Russell for technical help. This study was supported by grants from the Deutsche Forschungsgemeinschaft (to W.R. and B.K.F.), the Federal Ministry of Education and Research, Germany (to T.L.), the European Commission (to B.K.F.), BONFOR (to J.W.S., A.H.) and the National Institutes of Health (to Y.N.T., M.I.K. and G.S.).

Author Contributions W.R., T.L. and P.S. contributed equally to this work. W.R. performed microsurgery, left-ventricular catheterization and analysis of data. T.L. supervised and analysed *in vivo* electrophysiology. P.S. performed *in vivo* imaging experiments, analysis and immunohistochemistry. Y.N.T. was involved in the *in vivo* imaging experiments and immunohistochemistry. B.-R.C. was involved in optical imaging and analysis of data. M.B. was involved in morphometry, immunohistochemistry and establishment of cardiac fibroblasts. R.D. was involved in the analysis of *in vivo* imaging experiments. U.B. was involved in the preparation of skeletal myoblasts, and in morphometry and immunohistochemistry. S.-M.H. was involved in the optical-imaging experiments and analysis of data. T.B. was involved in microsurgery, morphometry, left ventricular catheterization and immunohistochemistry. J.V.M. was involved in mouse breeding, immunohistochemistry and western blotting of tissues from the Cx43-expressing mouse. A.H. was involved in the generation of the lentivirus constructs. S.R. was involved in mouse breeding and *in vivo* imaging experiments. B.D. was involved in generation of the Cx43-expressing mouse model. B.G. was involved in Langendorff perfusion and optical-imaging experiments. A.P. supervised lentiviral work. A.W. supervised the microsurgery. G.S. was involved in the optical-imaging experiments, their analysis and writing of the manuscript. J.W.S. performed electrophysiological experiments *in vivo* and analysed data. M.I.K. designed experiments, analysed *in vivo* imaging experiments and wrote the manuscript. B.K.F. initiated the study, designed experiments and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to B.K.F. (bernd.fleischmann@uni-bonn.de) or M.I.K. (mik7@cornell.edu).

METHODS

All animal experiments were performed in accordance with National Institutes of Health animal protection guidelines and were approved by the local authorities. **Preparation and characterization of donor cells.** Cardiac and skeletal muscle cells expressing EGFP under control of the human α -actin promoter were harvested from transgenic mice (HIM:OF1 and CD1 background)¹⁶ and transplanted into syngeneic wild-type mice. Skeletal muscle was harvested from hindlimb muscles of E18.5/E19.5 embryos and dissected; single cells were obtained by dissociation with collagenase and trypsin. Embryonic hearts (E14.5–E16.5) were enzymatically dissociated into single cells as reported previously²⁹. To assess the percentage of eCMs after dissociation, flow cytometry was performed. Cardiomyocytes in transgenic α -MHC-EGFP⁺ (ref. 32) embryonic hearts (E15.5) were found to be $46.8 \pm 5.6\%$ ($n = 3$) using flow cytometry.

SMs from the hindlimb and diaphragm of embryonic (E18.5/E19.5) and postnatal (day 1/day 2) Cx43tetOeGFP/MCKiTA (SM-Cx43⁺) transgenic mice were harvested and isolated as described above. For negative controls, skeletal muscle was harvested and processed in an identical fashion from monoallelic littermates (Cx43tetOeGFP/– and –/MCKiTA: SM-Cx43[–]). These cells were transplanted either directly after dissociation or after 24 h of culture. Cardiac myofibroblasts (cFib, E14.5–E16.5) were isolated, cultured and characterized in a similar manner to that reported previously³³. BM cells were harvested by BM aspiration by flushing the femur and tibia of 2–3-month-old wild-type CD1 mice with a 27-gauge needle (see also ref. 28).

Transplantation. Reproducible transmural cryolesions at the anterior-lateral left ventricular wall were generated in recipient mice (CD1, HIM:OF1, C57/Bl6, 10–12 weeks) with liquid-N₂-cooled copper probes 3 or 4 mm in diameter as reported previously^{28,29} and cells suspended in 5–6 μ l were injected intramyocardially immediately thereafter. SM-Cx43⁺ and SM-Cx43[–] cells (C57/Bl6 background) were transplanted into CD1 mice. For non-syngeneic transplantations, the recipients were immunosuppressed by using daily intraperitoneal injections of cyclosporin A (20 mg kg^{–1}; Novartis).

Left ventricular catheterization. Left ventricular catheterization was performed 2 weeks after the operation by a blinded investigator with a 1.4 french Millar Atrial catheter (Millar Instruments Inc.)²⁸.

Surface ECG and *in vivo* transvenous electrophysiological investigation. Mice (HIM:OF1 or CD1 background) were put under inhalative anaesthesia and a surface 6-lead ECG (lead III is shown in figures) was obtained for at least 3 min. R–R intervals, P-wave durations, PQ intervals, QRS durations and QT and JT intervals were measured by successive evaluation as described previously^{30,34,35}. A schematic illustration of the measured surface ECG parameters is given in Supplementary Fig. 1a. JT was defined as the interval from the S-wave meeting the isoelectrical line to the end of the T-wave. QT and JT intervals were also rate-corrected as reported previously³⁵. After amplification, all data were sampled at 4 kHz (Bard stamp amplifier; C.R. Bard Inc.). For transvenous electrophysiological investigation a 2 french octapolar mouse-electrophysiological catheter (CIBer Mouse; NuMed Inc.) was inserted into the right jugular vein and positioned in the right atrium and ventricle. Bipolar electrograms were obtained from adjacent electrode pairs. Twice pacing threshold rectangular stimulus pulses were administered by a multi-programmable stimulator (Model 5328; Medtronic). Evaluation of electrophysiological investigation parameters included sinus node recovery period (SNRP), Wenckebach periodicity (WBP) as well as atrial refractory periods (ARP), atrio-ventricular-nodal refractory periods (AVNRP) and ventricular refractory periods (VRP). Vulnerability to VT, characterized by atrio-ventricular dissociation, was tested by ventricular burst stimulations performed at a stimulation cycle length (S1S1-CL) starting at 50 ms with 10-ms stepwise reduction down to 10 ms (refs 30, 36). In addition, extrastimulus pacing was used with eight fixed-rate stimuli at S1S1-CL of 120, 110 and 80 ms, followed by up to three short-coupled extra beats with successive reduction of the coupling interval until VRP was reached. Stimulations were performed at twice pacing threshold. VT was defined as at least four ventricular beats; the mean VT duration was 3.4 ± 5 s ($n = 9$). For all mice, complete vulnerability testing was performed; ECG analysis was not performed in some sham-injected and eCM-engrafted mice.

***In vivo* imaging.** GCaMP2⁺ eCMs were obtained from transgenic embryos¹⁹ and transplanted into non-transgenic littermates or wild-type C57/Bl6 mice. Alternatively, eCMs were harvested from CD1 wild-type embryonic hearts (E14.5–E16.5) and transduced overnight *in vitro* with lentiviral vectors expressing GCaMP2 under the control of the cytomegalovirus (CMV) promoter (LVGCaMP2)³¹. In brief, LVGCaMP2 was cloned by replacing the EGFP in LV-GFP³¹ with the GCaMP2 complementary DNA. Lentiviral particles were prepared as described previously³¹. The cells were trypsinized and thoroughly washed after lentivirus transduction and injected (10^3 to 4×10^5 cells) into recipient CD1 mice. Imaging experiments were performed 7–16 days after

operation in ventilated and thoracotomized mice in which the heart was exposed and partly immobilized. Engrafted cells were revealed with a fluorescence macro-scope (OV100 or MVX10; Olympus). GCaMP2 was excited at 470 ± 20 nm, emitted light was collected at 525 ± 25 nm with a electron-multiplied, charge-coupled device camera (iXon 860–BI or iXion 885KS; Andor Technology). Data were acquired at frame rates of 67–128 Hz with the Andor Solis software. ECG was recorded by an EXT-02F amplifier (npi electronic) connected to a PowerLab 26T AD converter (AD Instruments). Data were processed with ImageJ software (NIH) and custom-written software (Labview 7.1; National Instruments). In some experiments, the concentration of isoflurane was increased to decrease the heart rate to less than 300 beats per minute. Some of the hearts were explanted and imaged while being perfused with a Langendorff system. Registration using the ‘turbo reg’ plugin of ImageJ³⁷ was used to eliminate translating and rotating movement. Pseudo-coloured images were generated after individual normalization of each pixel.

Optical mapping. For optical-mapping experiments, hearts (CD1 or C57/Bl6 background) were excised, perfused in a constant-flow Langendorff chamber and stained with di-4-ANEPPS or Rhod-2 as described previously²³. Fluorescence (di-4-ANEPPS: $\lambda_{ex} = 540 \pm 30$ nm, $\lambda_{em} > 640$ nm; Rhod-2: $\lambda_{ex} = 540 \pm 30$ nm, $\lambda_{em} = 585 \pm 20$ nm) was collected with a camera lens (50 mm focal length, $f 0.95$; Navitar), focused on a CMOS camera (100 pixels \times 100 pixels; Ultima-L SciMedia). Excitation light was delivered by epi-illumination with a custom-designed 300-W tungsten–halogen lamp (University of Pittsburgh machine and electronic shops). The CMOS camera (1×1 cm²) viewed the anterior surface of the heart, with each pixel recording signals from four to ten adjacent cardiomyocytes ($70 \times 70 \mu$ m² per pixel with a depth of field of about 130–140 μ m); the integration time was set to 1 ms (1,000 frames s^{–1}) to reduce motion blurring caused by rapid wave propagation. To measure the conduction velocity from the infarct region and its healthy surroundings, point stimulation (Ag⁺ electrode 200 μ m in diameter; 2 ms duration; 20% above threshold) was performed at a cycle length of 200 ms for at least 20 beats. Activation time-points were determined by dF/dt_{max} after applying a Savitzky–Golay smoothing filter (11 point width, third order), and isochronal maps of activation were generated as described previously³⁸. The local conduction velocity vector was calculated from the activation time-point at each pixel compared with that of its eight nearest neighbours^{19,23,25}. The rise time of action potentials was measured from fluorescence signals that showed a signal-to-noise ratio of more than about 20:1, at a sampling rate of 5,000 frames s^{–1} and a spatial resolution of $100 \times 100 \mu$ m² per pixel. The rise time was measured from 10% to 90% of action-potential amplitude from at least five consecutive action potentials and low-pass filtered with an averaging kernel of 1 ms. Border zones³⁹ were identified from activation maps where slow propagation occurred and from autofluorescence images.

Histology and Immunohistochemistry. After electrophysiological investigation, hearts were harvested, imaged with a fluorescence stereomicroscope (Leica MZ 16F; Leica Microsystems) and a ProgRes C10+ camera (Jenoptik), arrested in diastole by coronary perfusion with cardioplegic solution at stable haemodynamic pressure, and fixed with 4% paraformaldehyde by perfusion. The hearts were cryopreserved and cut into tissue slices 8 or 20 μ m thick (for morphometry and immunohistochemistry) and 20 μ m thick (for histology)¹⁵. The extent of myocardial lesions in the different groups of animals was determined by staining with Sirius red (Aldrich Chemical Company). Immunostaining with anti-EGFP (1:50 dilution from Santa Cruz Biotechnology, or 1:20 dilution from Chemicon) was used to corroborate EGFP or GCaMP2 expression. To evaluate cross-striation, primary antibodies against sarcomeric α -actinin were used. The differentiation of engrafted SMs was confirmed by using a mouse monoclonal antibody directed against nebulin (1:10 dilution). Cx43 immunostaining was performed with a commercially available antibody (1:400 dilution; Bio Trend) or an antibody provided by K.W.⁴⁰. All primary antibodies were revealed by appropriate secondary Cy3-conjugated and Cy5-conjugated donkey antibodies (1:400–1:1,000 dilution; Dianova). Nuclei were stained blue with Hoechst dye. Samples were imaged with a Zeiss Axiovert 200 microscope equipped with an ApoTome and AxioCam MRm; images were acquired with the Zeiss software AxioVision.

Statistical analysis. Statistical analysis of the electrophysiological investigation data was performed with a multivariate one-way analysis of variance with post-hoc subgroup testing when appropriate (Tukey–Kramer multiple comparisons test). Non-parametric variables were evaluated by using Kruskal–Wallis testing with Dunn post-hoc testing when appropriate. Discrete variables were analysed by two-sided Fisher’s exact test. Statistical analysis of the other data was performed by Student’s *t*-test. $P \leq 0.05$ was regarded as statistically significant; unless stated otherwise, errors are given as s.d.

32. Rubart, M. *et al.* Physiological coupling of donor and host cardiomyocytes after cellular transplantation. *Circ. Res.* **92**, 1217–1224 (2003).

33. Miragoli, M., Gaudesius, G. & Rohr, S. Electrotonic modulation of cardiac impulse conduction by myofibroblasts. *Circ. Res.* **98**, 801–810 (2006).
34. Berul, C. I., Aronovitz, M. J., Wang, P. J. & Mendelsohn, M. E. *In vivo* cardiac electrophysiology studies in the mouse. *Circulation* **94**, 2641–2648 (1996).
35. Mitchell, G. F., Jeron, A. & Koren, G. Measurement of heart rate and Q-T interval in the conscious mouse. *Basic Res. Cardiol.* **274**, H747–H751 (1998).
36. Schrickel, J. W. *et al.* Induction of atrial fibrillation in mice by rapid transesophageal atrial pacing. *Basic Res. Cardiol.* **97**, 452–460 (2002).
37. Thevenaz, P., Ruttimann, U. E. & Unser, M. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* **7**, 27–41 (1998).
38. Choi, B. R. & Salama, G. Simultaneous maps of optical action potentials and calcium transients in guinea-pig hearts: mechanisms underlying concordant alternans. *J. Physiol. (Lond.)* **529**, 171–188 (2000).
39. Gottlieb, G. J., Kubo, S. H. & Alonso, D. R. Ultrastructural characterization of the border zone surrounding early experimental myocardial infarcts in dogs. *Am. J. Pathol.* **103**, 292–303 (1981).
40. Traub, O. *et al.* Immunochemical and electrophysiological characterization of murine connexin40 and -43 in mouse tissues and transfected human cells. *Eur. J. Cell Biol.* **64**, 101–112 (1994).

Bv8 regulates myeloid-cell-dependent tumour angiogenesis

Farbod Shojaei¹, Xiumin Wu¹, Cuiling Zhong¹, Lanlan Yu¹, Xiao-Huan Liang¹, Jenny Yao¹, Dominique Blanchard¹, Carlos Bais¹, Franklin V. Peale¹, Nicholas van Bruggen¹, Calvin Ho¹, Jed Ross¹, Martha Tan¹, Richard A. D. Carano¹, Y. Gloria Meng¹ & Napoleone Ferrara¹

Bone-marrow-derived cells facilitate tumour angiogenesis, but the molecular mechanisms of this facilitation are incompletely understood. We have previously shown that the related EG-VEGF and Bv8 proteins, also known as prokineticin 1 (Prok1) and prokineticin 2 (Prok2), promote both tissue-specific angiogenesis and haematopoietic cell mobilization. Unlike EG-VEGF, Bv8 is expressed in the bone marrow. Here we show that implantation of tumour cells in mice resulted in upregulation of Bv8 in CD11b⁺Gr1⁺ myeloid cells. We identified granulocyte colony-stimulating factor as a major positive regulator of Bv8 expression. Anti-Bv8 antibodies reduced CD11b⁺Gr1⁺ cell mobilization elicited by granulocyte colony-stimulating factor. Adenoviral delivery of Bv8 into tumours was shown to promote angiogenesis. Anti-Bv8 antibodies inhibited growth of several tumours in mice and suppressed angiogenesis. Anti-Bv8 treatment also reduced CD11b⁺Gr1⁺ cells, both in peripheral blood and in tumours. The effects of anti-Bv8 antibodies were additive to those of anti-Vegf antibodies or cytotoxic chemotherapy. Thus, Bv8 modulates mobilization of CD11b⁺Gr1⁺ cells from the bone marrow during tumour development and also promotes angiogenesis locally.

Angiogenesis has an important role in tumour progression and metastasis, and anti-angiogenesis represents a clinically validated anti-cancer strategy^{1–3}. Although tumour cells were traditionally considered to be the main source of angiogenic factors⁴, it has become apparent that untransformed tumour-infiltrating cells, such as fibroblasts^{5,6}, immune cells⁷, endothelial progenitors^{8,9} and different subpopulations of myeloid cells, may have a role in tumour angiogenesis^{10–12} and in vascular endothelial growth factor (Vegf, also known as Vegfa)-induced neovascularization¹³. CD11b⁺Gr1⁺ myeloid cells have recently been shown to mediate refractoriness to anti-Vegf therapy in some tumour models¹⁴. A role for neutrophils in the angiogenic switch in a transgenic tumour model has been described¹⁵. Myeloid cells may locally secrete angiogenic factors or matrix metalloproteinase 9 (refs 11, 15 and 16), which may in turn increase Vegf bioavailability¹⁷.

Bv8 and EG-VEGF are two highly related secreted proteins that belong to a larger class of peptides defined by a five-disulphide-bridge motif called a colipase fold^{18–21}. Bv8 was identified initially as a secreted protein from the skin of the frog *Bombina variegata*, which induced hyperalgesia and gastrointestinal motility²¹. Later on, the human orthologue of this highly conserved protein was shown to have the same activities¹⁸. Bv8 and EG-VEGF bind two highly related G-protein-coupled receptors (GPCRs), EG-VEGFR/PKR1 and EG-VEGFR/PKR2^{22,23}. We characterized EG-VEGF and Bv8 as mitogens selective for specific endothelial cell types^{19,20,24}. Furthermore, Bv8 or EG-VEGF induced haematopoietic cell mobilization *in vivo* and stimulated production of granulocytic and monocytic colonies *in vitro*^{25,26}. These activities, combined with expression of Bv8 (but not of EG-VEGF) in the bone marrow²⁵, suggest that Bv8 may be an especially interesting candidate for a mediator of inflammatory-cell-dependent angiogenesis.

G-CSF regulates Bv8 expression in bone marrow

To determine whether Bv8 expression in the bone marrow is affected by tumour growth at a distant site, we transplanted A673 or HM7

tumour cells in immunodeficient mice. As illustrated in Fig. 1a, enzyme-linked immunosorbent assay (ELISA) was used to demonstrate that tumour implantation resulted in a significant increase in Bv8 levels in the bone marrow mononuclear cells (BMMNCs) compared to empty Matrigel. To characterize the Bv8 protein, we subjected lysates of mouse BMMNCs to heparin–Sepharose affinity chromatography. As illustrated in Supplementary Fig. 1a, Bv8 bound to the column and was eluted in the presence of ~0.4 M NaCl. Western blot analysis confirmed the presence of the expected ~9 kDa band in the immunoreactive fractions (Supplementary Fig. 1b).

BMMNCs comprise several subsets of cells, mainly of myeloid and lymphoid lineages. To elucidate which subset of BMMNCs is enriched in Bv8, we implanted several tumour cell lines in mice, including A673, Calu6, HM7, HPAC and Jurkat cells. Bv8 was highly expressed in CD11b⁺Gr1⁺ cells (consisting primarily of neutrophils, but also including cells of the macrophage lineage^{11,27,28}) compared to CD11b⁺Gr1[−] cells (mostly non-myeloid subset; Fig. 1b).

To identify potential regulators of Bv8 expression, we tested the ability of a panel of cytokines and chemokines to induce Bv8 expression in unsorted BMMNCs (Supplementary Fig. 2a). Although most of the cytokines had no significant effect, the presence of granulocyte colony-stimulating factor (G-CSF, also known as Csf3; 10 ng ml^{−1}) resulted in a marked increase in Bv8 expression (>27-fold; Supplementary Fig. 2a). Analysis of different subsets of BMMNCs revealed that the presence of G-CSF results in Bv8 upregulation more than 100-fold above background levels in purified CD11b⁺Gr1⁺ cells (Fig. 1c). A substantially lower upregulation was detected in whole BMMNCs or in the CD11b⁺Gr1[−] fraction. Surprisingly, granulocyte-macrophage CSF (GM-CSF; also known as Csf2) had no effect on Bv8 expression, suggesting a high degree of specificity in Bv8 regulation (Fig. 1c). Interleukin 6 (IL-6) and the chemokine Sdf1 (also known as Cxcl12), which did not show any significant stimulation when tested on unsorted BMMNCs, resulted in a 2–3-fold upregulation of Bv8 on

¹Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA.

purified $CD11b^+Gr1^+$ cells (Fig. 1c). We also tested whether lower, more physiologically relevant concentrations of G-CSF may induce *Bv8* expression in $CD11b^+Gr1^+$ cells under normoxic or hypoxic conditions²⁰. As shown in Supplementary Fig. 2b, as little as 20 pg ml^{-1} G-CSF resulted in ~ 10 -fold stimulation of *Bv8* expression under hypoxic conditions. This stimulation was significantly higher than that detected under normoxic conditions (~ 5 -fold). Furthermore, *in vivo* administration of recombinant G-CSF to Balb/c mice resulted in a time- and dose-dependent increase in the levels of *Bv8* protein in the BMMNCs (Supplementary Fig. 1c) and in the serum (Supplementary Fig. 2d), coincident with an increase in peripheral blood neutrophils (Supplementary Fig. 2e). Similar results

were obtained in Balb/c nude mice (data not shown). As early as 24 h after G-CSF administration, the *Bv8* levels in the bone marrow increased ~ 30 -fold above background levels. Serum levels were also markedly increased. However, bone marrow and serum *Bv8* returned to near baseline levels within 48 h after discontinuing G-CSF treatment (Supplementary Fig. 2d, e). G-CSF administration had no effect on *Bv8* levels in kidney, brain and liver (data not shown).

G-CSF is a principal regulator of granulopoiesis²⁹ and also has a key role in neutrophil mobilization from the bone marrow in response to a variety of environmental stresses³⁰. G-CSF is secreted by several cell types, including endothelial cells and fibroblasts³⁰, and, together with other cytokines including IL-6 and Sdf1, is constitutively expressed by tumour and/or stromal cells in malignant tumours³¹. Analysis of conditioned media isolated from tumour cells or tumour-associated fibroblasts showed that G-CSF, IL-6 and Sdf1 were detectable in both compartments (Supplementary Table 1). To define the potential role of G-CSF in regulating *Bv8* expression within the tumour microenvironment, cultured BMMNCs were treated with aliquots of lysates from HM7 tumours pre-incubated with anti-G-CSF or control antibodies (Fig. 1d). Taqman analysis demonstrated a significant dose-dependent reduction in *Bv8* expression in BMMNCs treated with anti-G-CSF compared to the control immunoglobulin G (IgG)-treated wells. *In vivo* studies confirmed a key role of G-CSF in regulating *Bv8* expression. *Bv8* protein was significantly reduced in the bone marrow of non-tumour-bearing mice treated with anti-G-CSF compared to controls (Supplementary Fig. 2f). In addition, anti-G-CSF antibody virtually abolished the peak in *Bv8* protein occurring in the bone marrow shortly after tumour implantation (Fig. 1e). Anti-G-CSF treatment also resulted in a significant reduction in the frequency of circulating as well as bone marrow $CD11b^+Gr1^+$ cells in non-tumour- and tumour-bearing mice (Supplementary Fig. 2g–j). Therefore, although we do not rule out the involvement of additional factors, *Bv8* expression is crucially dependent on G-CSF, both *in vitro* and *in vivo*. Thus, upregulation of *Bv8* in the bone marrow mediated by G-CSF (and/or another cytokine) may contribute to mobilization of myeloid cells. Subsequent homing of these cells in the tumours may be regulated by additional cytokines and also potentially by *Bv8* secreted by tumour-associated myeloid cells.

We wished to determine whether *Bv8* contributes to mobilization of neutrophils induced by recombinant G-CSF. A sub-maximal dose of G-CSF ($2\text{ }\mu\text{g}$) induced a significant mobilization of $CD11b^+Gr1^+$ cells to the peripheral blood of mice within 6 h of administration (Fig. 1f). The effect of G-CSF was completely blocked by an anti-G-CSF antibody. Anti-*Bv8* antibodies (hereafter, anti-*Bv8*) also inhibited G-CSF-mediated mobilization of $CD11b^+Gr1^+$ cells. However, anti-*Bv8* had little effect on the mobilization induced by a maximal dose of G-CSF ($10\text{ }\mu\text{g}$), at least at the dose tested. Therefore, *Bv8* may function to modulate or amplify neutrophil mobilization stimulated by G-CSF, by paracrine or autocrine mechanisms.

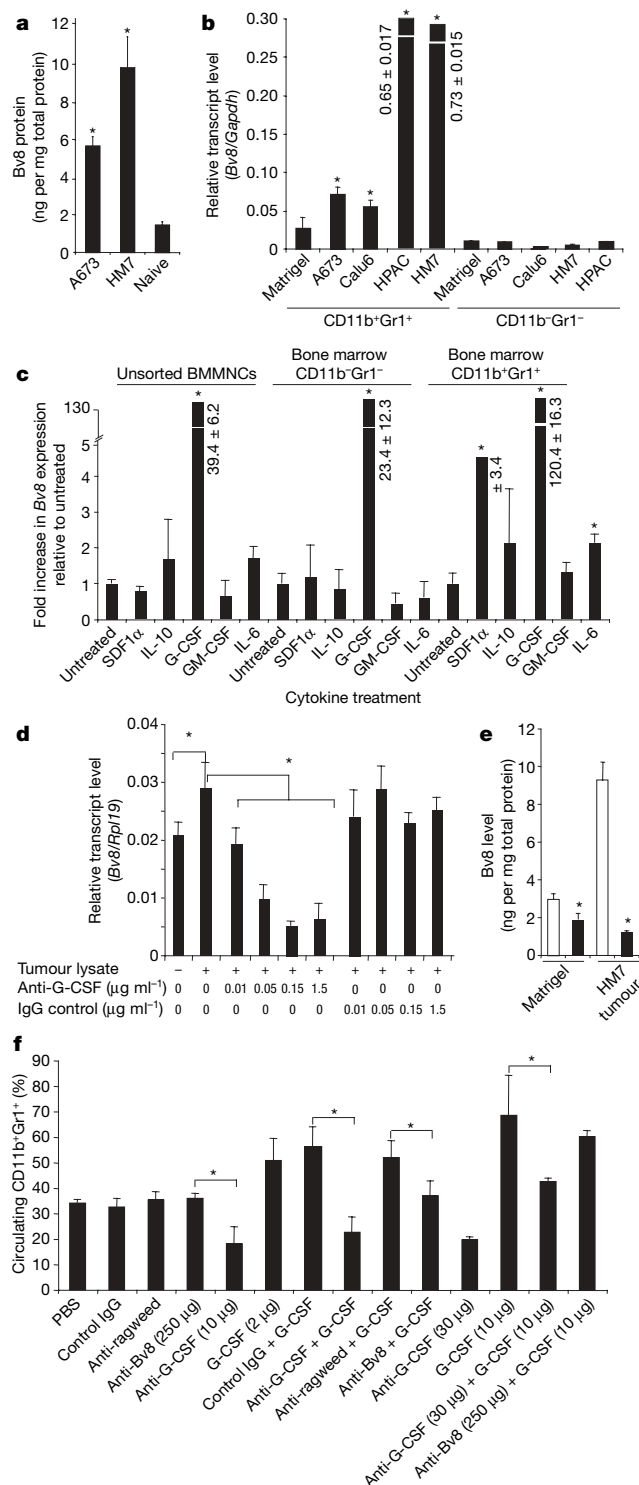


Figure 1 | Regulation of *Bv8* expression *in vitro* and *in vivo*. **a**, Beige nude mice were implanted with A673 ($n = 6$) or HM7 ($n = 6$) cells, and *Bv8* was measured in the bone marrow by ELISA. **b**, Beige nude mice were implanted ($n = 5$) with A673, HM7, HPAC, Calu6 or Jurkat cells. *Bv8* expression in BMMNC $CD11b^+Gr1^+$ and $CD11b^-Gr1^-$ subsets was determined by Taqman. *Gapdh*, glyceraldehyde-3-phosphate dehydrogenase. **c**, Whole BMMNCs, $CD11b^+Gr1^+$ and $CD11b^-Gr1^-$ populations were treated with Sdf1 α , IL-6, IL-10, G-CSF or GM-CSF, and *Bv8* expression was measured ($n = 4$). **d**, BMMNCs were treated with aliquots of HM7 tumour lysates, pre-incubated with various concentrations of anti-G-CSF or control IgG. *Bv8* expression was monitored by Taqman ($n = 4$). *Rpl19*, ribosomal protein L19. **e**, Anti-G-CSF reduces the level of *Bv8* in the bone marrow of HM7-tumour-bearing mice ($n = 6$). Clear bars, control IgG; filled bars, anti-G-CSF. **f**, *Bv8* has a role in bone marrow cell mobilization induced by G-CSF ($n = 5$). See Methods for details. All mice were bled 6 h after final injection, and the frequency of $CD11b^+Gr1^+$ cells was determined. Single asterisk, $P < 0.05$; error bars represent standard deviation (s.d.).

Anti-Bv8 antibodies inhibit tumour growth *in vivo*

To define the localization of Bv8 in tumours, we sub-fractionated the cell populations in A673, Calu6, HM7, HPAC and Jurkat tumours into CD11b⁺ and CD11b⁻ fractions. Taqman analysis showed a significant upregulation of mouse *Bv8* transcript in the CD11b⁺ compartment compared to the negative fraction (Fig. 2a). However, using human *BV8* primers, polymerase chain reaction (PCR) did not identify any human *BV8* in either population (Fig. 2a), that is, tumour-associated CD11b⁺ and CD11b⁻, suggesting that the mouse stroma, particularly myeloid cells, is the main source of Bv8 in all tumours tested. In agreement with these findings, none of the tumour cell lines tested produced detectable levels of BV8 protein *in vitro* by ELISA (data not shown).

We tested whether treatment with anti-Bv8, initiated shortly after tumour cell inoculation, may affect the growth of several tumour cell lines, including A673, HM7, Calu6 and Jurkat tumours, in Balb/c nude mice. As illustrated in Fig. 2b–e, administration of anti-Bv8 resulted in a significant decrease in tumour volume and weight compared to control-treated animals in all tumour models examined. In some models, the growth inhibition elicited by anti-Bv8 approached that elicited by anti-Vegf monoclonal antibodies (hereafter, anti-Vegf) G6-31 or B20, which block both mouse Vegf and human VEGF³² (Fig. 2b). The tumour-implantation experiments shown were performed in Balb/c nude mice, but similar results were obtained in

Beige nude mice (data not shown). In a similar manner to human xenografts, anti-Bv8, alone or in combination with anti-Vegf, reduced the growth of the TIB42 mouse tumour (Fig. 2f). Discontinuation of the treatment resulted in rapid tumour growth in mice bearing A673 (Supplementary Fig. 3g) or HM7 (Supplementary Fig. 3h) tumours, concomitant with an increase in the number of tumour-infiltrating CD11b⁺Gr1⁺ cells (Supplementary Fig. 3i, j).

Anti-Bv8 treatment reduces myeloid cell mobilization

We performed a series of *in vitro* and *in vivo* studies to characterize the role of Bv8 in the haematopoietic system (see Supplementary Information and Supplementary Figs 3–7). To study the role of myeloid cells during tumour growth, we investigated the kinetics of CD11b⁺Gr1⁺ cells in the bone marrow, peripheral blood and tumours, at different time points, in mice implanted with A673 cells (Supplementary Fig. 5). Although there was no significant difference in the frequency of bone marrow CD11b⁺Gr1⁺ cells (Supplementary Fig. 5a), we found a significant reduction in the numbers (data not shown) and frequency of CD11b⁺Gr1⁺ cells in the peripheral blood in anti-Bv8 compared to control-treated mice at all time points tested (Supplementary Fig. 5b). Furthermore, anti-Bv8 elicited a significant reduction in the number of tumour-associated CD11b⁺Gr1⁺ cells at several time points in the same model (Supplementary Fig. 5c). Using flow cytometry (representative fluorescence-activated cell sorting

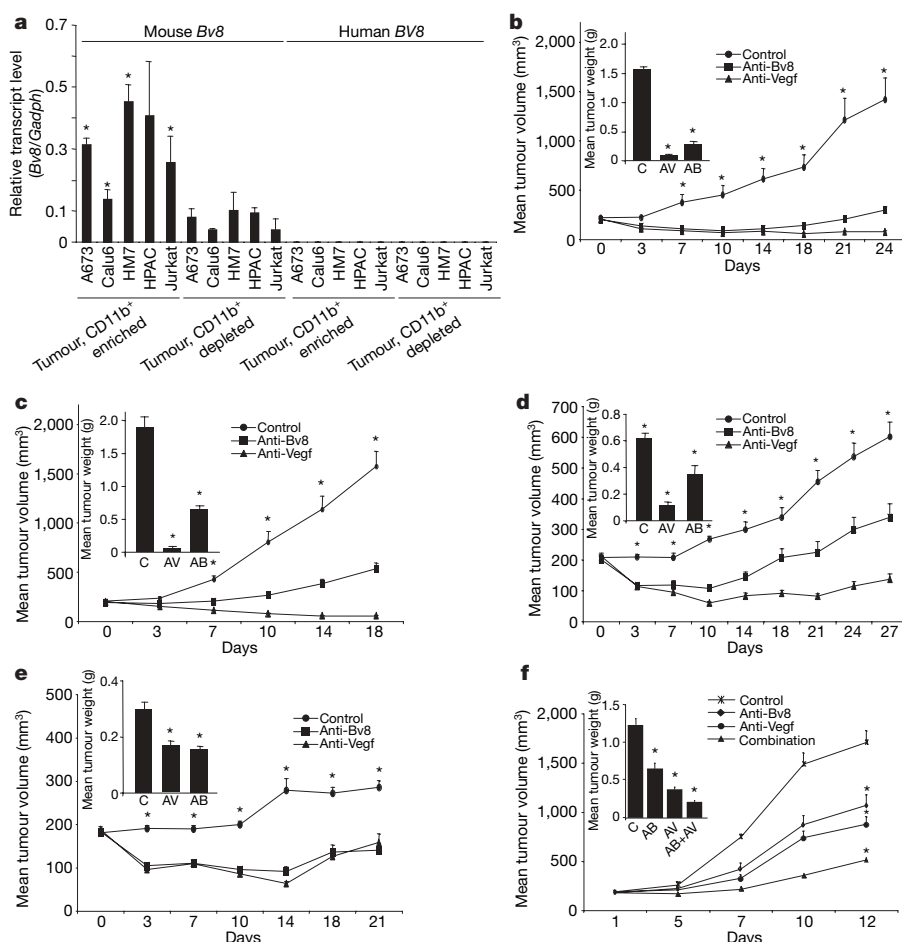


Figure 2 | Expression of Bv8 in tumour-associated myeloid cells, and effects of anti-Bv8 antibodies on tumour growth. **a**, CD11b⁺ cells are the main source of Bv8 in the tumours. Beige nude mice ($n = 5$) were implanted with A673, Calu6, HM7, HPAC or Jurkat cells and were killed ten days after tumour cell transplantation. Populations of cells enriched for CD11b⁺ were isolated and the expression of mouse *Bv8* or human *BV8* was analysed. **b–e**, Balb/c nude mice ($n = 10$) were implanted with A673 (**b**), HM7 (**c**), HPAC (**d**) or Jurkat (**e**) cells and were treated as described. Asterisks

indicate significant difference in anti-Bv8 or anti-Vegf compared to control-treated groups ($P < 0.05$). C, control; AB, anti-Bv8; AV, anti-Vegf. Insets show tumour weights. **f**, Anti-Bv8 and anti-Vegf have additive effects to inhibit tumour growth. Beige nude mice ($n = 10$) were implanted with TIB42 cells and were treated with control, anti-Bv8, anti-Vegf or a combination (anti-Bv8 plus anti-Vegf). The inset shows the terminal tumour weights in all treatments. Error bars represent s.e.m.

(FACS) profiles are shown in Supplementary Fig. 6), the kinetics of $CD11b^+Gr1^+$ populations were also investigated in the peripheral blood, tumours, bone marrow and spleen of mice implanted with Calu6, HM7, HPAC and Jurkat cells (Fig. 3a, b and Supplementary Fig. 7a, b). Consistent with the time course study in the A673 model, treatment with anti-Bv8 resulted in a significant decrease in the frequency of peripheral blood (Fig. 3a) and tumour-infiltrating (Fig. 3b) $CD11b^+Gr1^+$ cells in all tumour models examined. These findings indicate that Bv8 regulates mobilization and, potentially, homing of $CD11b^+Gr1^+$ cells to the tumours. In addition, neutrophils (mainly identified by the expression of Gr1; ref. 33) seem to be the main population affected by anti-Bv8 treatment.

Previous studies have shown that transplantation of myeloid cells, including $CD11b^+Gr1^+$ cells, enhances tumour growth, whereas their depletion may reduce it^{11,12,14,15}. To assess directly the role of myeloid cells in Bv8-regulated tumour growth, we isolated bone marrow $CD11b^+Gr1^+$ cells from mice bearing A673 or HM7 tumours seven days after tumour implantation, and injected them into the tumours in mice engrafted with the same tumours for seven days, with or without anti-Bv8 treatment. This resulted in a more rapid tumour growth in anti-Bv8-treated animals (Supplementary Fig. 5d, e). Therefore, an excess of $CD11b^+Gr1^+$ cells may override the tumour growth inhibition elicited by anti-Bv8 treatment. In addition, we characterized kinetics of $CD11b^+Gr1^+$ cells in bone marrow and spleen (see Supplementary Information).

Bv8 regulates tumour angiogenesis

To determine whether Bv8 may also affect the tumour vasculature directly, we tested recombinant Bv8 on mouse tumour-associated endothelial cells and found that it induces mitogen-activated protein (MAP) kinase phosphorylation and tube formation (Supplementary Fig. 8a–c). However, it had no mitogenic effects on several tumour cell lines (Supplementary Fig. 8d). We next delivered intra-tumourally recombinant adenovirus mouse Bv8 (*Av-Bv8*) in HM7-tumour-bearing mice (10^7 plaque-forming units, p.f.u.). *Av-LacZ* and *Av-Vegf* served as negative and positive controls, respectively.

Compared to adding control *Av-LacZ*, *Av-Bv8* administration resulted in an increase in tumour volumes, comparable to that induced by *Av-Vegf* (Fig. 4a). Consistent with the *in vitro* observations, administration of *Av-Bv8* resulted in enhanced mobilization of $CD11b^+Gr1^+$ compared to *Av-LacZ* and *Av-Vegf* (Fig. 4b). Higher titres (10^9 p.f.u.) of *Av-Bv8* also enhanced tumour growth and resulted in higher mobilization of $CD11b^+Gr1^+$ cells in the peripheral blood and in the tumour (data not shown).

To assess the tumour vasculature, we used X-ray micro-computed tomography (micro-CT)^{34–36}. Micro-CT analysis demonstrated that *Av-Bv8* and *Av-Vegf* had very similar effects, because both resulted in increases ($P < 0.05$) in vascular volume (Fig. 4c) compared to the *Av-LacZ* group. Representative images of the whole tumour mass from each treatment group are shown in Fig. 4e. The surface renderings of the extracted vascular network (red) and tumour (grey) are generated by the image-processing algorithm that defines the volumetric regions used in the analysis. In addition, MECA-32 immunostaining confirmed a significant increase in vascular surface areas in HM7 tumours after *Av-Bv8* administration relative to control *Av-LacZ* (Supplementary Fig. 9a).

To investigate further the role of Bv8 in tumour angiogenesis, we analysed the tumour vasculature in HM7 tumours treated with anti-Bv8, anti-Vegf or control antibodies by micro-CT. In agreement with the experiments illustrated in Fig. 2, anti-Bv8 treatment resulted in a significant reduction in tumour volume (Fig. 4f) and circulating $CD11b^+Gr1^+$ cells (Fig. 4g) compared to control. We found significant reductions in vascular volume in both the anti-Bv8 and anti-Vegf groups relative to control-treated tumours (Fig. 4h). Blood vessel density (vascular volume/tumour volume, VV/TV) was also reduced significantly in anti-Bv8 and anti-Vegf groups (Fig. 4i). Micro-CT data support the hypothesis that inhibition of tumour growth in anti-Bv8-treated mice is largely a consequence of suppression of tumour angiogenesis. A representative image of the entire tumour mass is shown in Fig. 4j. Therefore, gain- and loss-of-function approaches support the hypothesis that Bv8 promotes tumour growth primarily through induction of angiogenesis.

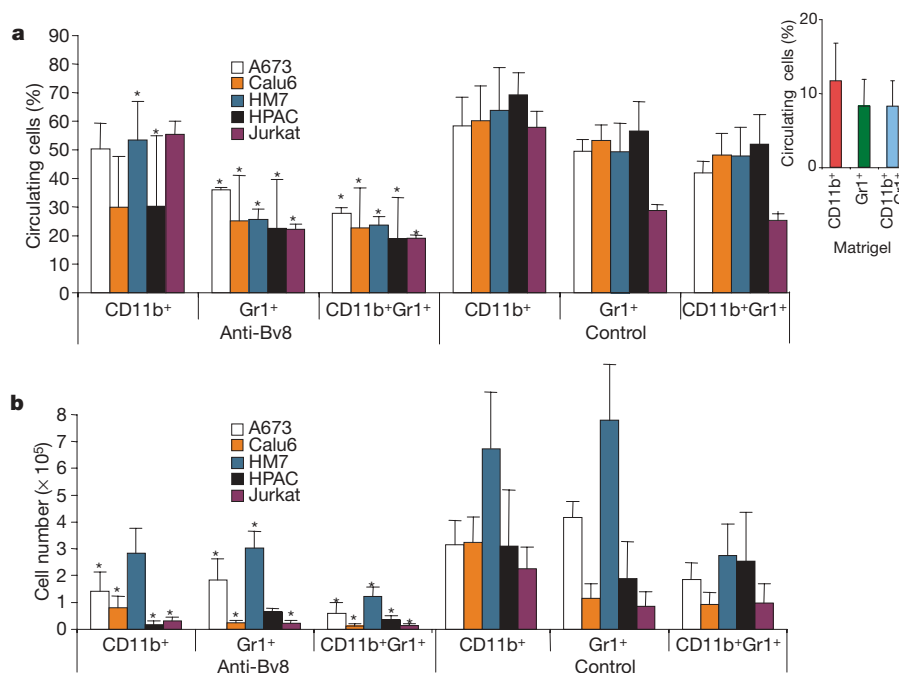


Figure 3 | Anti-Bv8 treatment reduces $CD11b^+Gr1^+$ cells in the peripheral blood and the tumours. **a, b**, Nude mice ($n = 5$) were implanted with A673, Calu6, HM7, HPAC or Jurkat cells and were treated with anti-Bv8 or control monoclonal antibodies, as described. Mice were analysed ten days after tumour cell inoculation and the frequency of $CD11b^+$, $Gr1^+$ and $CD11b^+Gr1^+$ cells was measured in peripheral blood (**a**) and in tumours

(**b**). The inset in **a** shows the frequency of circulating $CD11b^+$, $Gr1^+$ and $CD11b^+Gr1^+$ cells in Matrigel-implanted mice. Error bars represent s.d. Asterisks denote significant differences ($P < 0.05$) when comparing $CD11b^+$, $Gr1^+$ and $CD11b^+Gr1^+$ cells in each tumour in anti-Bv8 treated mice compared to control treated animals.

Histological examinations were also consistent with a role of Bv8 in tumour angiogenesis (Supplementary Fig. 9b). Analysis of endothelial cells in Jurkat tumours indicated that, similar to addition of anti-Vegf, administration of anti-Bv8 inhibited tumour vascularization.

Anti-Bv8 is additive to anti-Vegf or chemotherapy

We measured Bv8 protein levels in BMMNCs, sera, spleens and tumours in mice bearing HM7 tumours and treated with control or anti-Vegf (Supplementary Fig. 10a–d). Bv8 levels in control-treated mice showed a peak shortly after tumour implantation in bone marrow, peripheral blood and spleen. However, anti-Vegf-treated mice showed minimal Bv8 expression at such early stages, possibly owing to efficient tumour suppression elicited by the treatment. However, at later time points, coincident with the beginning of

Vegf-independent tumour growth, Bv8 levels were increased significantly in anti-Vegf-treated mice, especially in peripheral blood, spleen and tumours (Supplementary Fig. 10). In agreement with these findings, a large infiltration of Gr1⁺ cells was observed in necrotic areas of A673 and HM7 tumours that were treated with anti-Vegf for 15 or 21 days (Supplementary Fig. 11). A possible explanation is that long-term hypoxia and/or tumour necrosis elicited by anti-Vegf triggers neutrophil recruitment.

Therefore, anti-Bv8 treatment might be most effective when combined with anti-Vegf. To test this hypothesis, we implanted mice with HM7 (Fig. 5a) or A673 (Fig. 5b) cells and initiated treatments when tumours had reached 300–400 mm³. Consistent with the presence of lower intra-tumoural Bv8 levels, anti-Bv8 had a smaller effect on tumour growth inhibition in both HM7 and A673 tumours compared to early-stage treatment (Fig. 2). Anti-Vegf provided a more complete inhibition, but tumours eventually escaped. However, combination of anti-Vegf and anti-Bv8 significantly ($P < 0.05$) inhibited tumour growth compared to each monotherapy. Likewise, the combination therapy resulted in a significant reduction in tumour volumes and weight in TIB42 (Fig. 2f) and

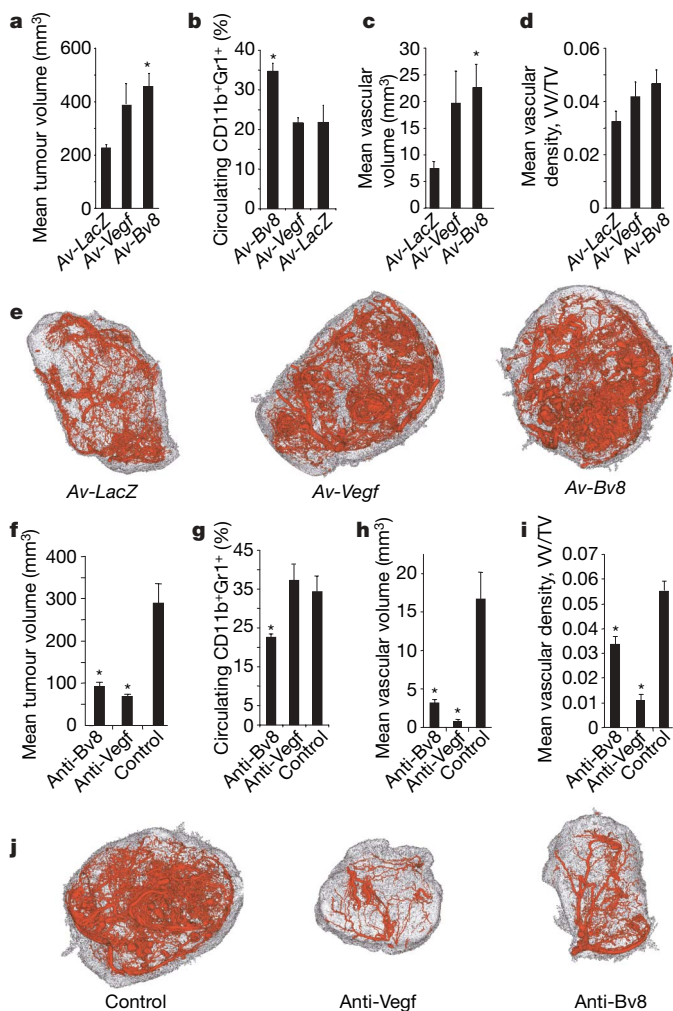


Figure 4 | Bv8 regulates tumour angiogenesis. HM7-tumour-bearing mice were injected with 10⁷ p.f.u. of Av-Bv8, Av-Vegf or Av-LacZ (each group, $n = 10$) five days after cell inoculation. **a**, Shown is the terminal tumour volume measurement in Av-Bv8 and Av-Vegf compared to Av-LacZ group. **b**, FACS data show the frequency of circulating CD11b⁺Gr1⁺ cells in the treatment groups. **c**, **d**, Micro-CT analysis revealed increased vascular volume (**c**) and vascular density (**d**) in Av-Bv8- and Av-Vegf-injected tumours. **e**, Representative images of vascular networks and tumours (shown in red and grey, respectively) from all treatment groups. **f**, Nude mice ($n = 10$) were implanted with HM7 cells and were treated with monoclonal antibodies. Data represent tumour volume measurement at day ten. **g**, FACS data represent the percentage of circulating CD11b⁺Gr1⁺ cells in treatment groups. **h**, **i**, The micro-CT analysis showed a significant decrease in vascular volume (**h**) and vascular density (**i**) in anti-Bv8- versus control-treated mice. **j**, Representative images of tumours from the three treatment groups. Error bars represent s.e.m. Asterisks denote significant differences ($P < 0.05$) when compared to controls.

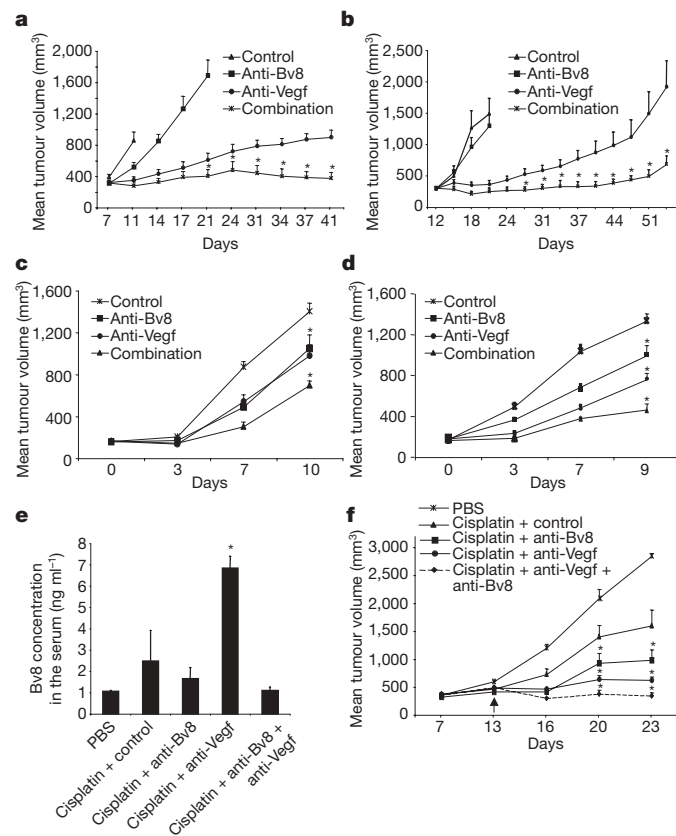


Figure 5 | Anti-Bv8 has additive effects with anti-Vegf or cytotoxic chemotherapy. **a**, **b**, Nude mice ($n = 10$) were implanted with HM7 (**a**) or A673 (**b**) tumour cells, and the indicated treatments were initiated when tumours reached 300–400 mm³. **c**, **d**, Nude mice ($n = 10$) and C57Bl/6 ($n = 10$) mice were implanted with EL4 cells and were treated with control, anti-Bv8, anti-Vegf or the combination. Asterisk indicates significant difference in tumour volume between combination therapy (anti-Bv8 plus anti-Vegf) versus each monotherapy. **e**, Cisplatin and anti-Vegf treatment increase Bv8 concentration in the serum. Nude mice ($n = 10$) were implanted with A673 cells and were treated with several agents in monotherapy or combination therapy as indicated. **f**, Chemotherapy plus anti-Vegf and anti-Bv8 effectively suppresses tumour growth in A673 tumours. Nude mice ($n = 10$) were injected with A673 cells. Treatments were initiated 13 days after tumour cell inoculation. Asterisk indicates significant difference between combination therapies versus cisplatin plus control. Error bars represent s.e.m.

EL4 mouse lymphomas (Fig. 5c), both of which are refractory to anti-Vegf¹⁴. Therefore, anti-Bv8 may be useful as a component of a combinatorial therapy for tumours that are refractory to anti-Vegf. To verify that the effects of anti-Bv8 are not limited to immunodeficient mice, we implanted the mouse anti-Vegf-resistant EL4 cell line in both immunodeficient and immunocompetent mice and tested the effects of anti-Bv8 and anti-Vegf monotherapy, or the combination. As illustrated in Fig. 5c, d, the effects of such treatments were almost indistinguishable in the two strains. These findings indicate that anti-Bv8 may suppress tumour growth even in the presence of an intact immune system.

Cytotoxic agents are known to result in mobilization of haematopoietic cells from the bone marrow³⁷. Furthermore, chemotherapy-induced tumour necrosis may result in release of chemokines such as G-CSF, followed by increased neutrophil production³⁸. Therefore, we sought to investigate whether treatment with cytotoxic agents, alone or in combination with anti-Vegf, affects the efficacy of anti-Bv8 treatment. Mice implanted with A673 cells were then treated with a major cytotoxic agent such as cisplatin, alone or in combination with anti-Bv8, anti-Vegf or with the two antibody treatments. Serum Bv8 was significantly ($P < 0.05$) increased in mice treated with cisplatin, alone or in combination with anti-Vegf (Fig. 5e). Both anti-Bv8 and anti-Vegf enhanced the anti-tumour activity of cisplatin. However, the combination of cisplatin plus anti-Vegf and anti-Bv8 resulted in almost complete inhibition of tumour growth in A673 cells ($P < 0.05$; Fig. 5f). Therefore, anti-Bv8 treatment could be used in combination with anti-Vegf and/or cytotoxic agents.

Discussion

Several activities have been associated with Bv8 signalling, including modulation of nociception²¹, gastrointestinal tract motility^{18,21}, circadian rhythms³⁹, olfactory bulb neurogenesis and survival of GnRH (gonadotropin-releasing hormone) neurons⁴⁰. Here we show that Bv8 has a highly specialized role in the bone marrow, in concert with G-CSF (see Supplementary Fig. 12 for a model), further emphasizing the versatility of this protein.

Growing evidence suggests that targeting myeloid cells may represent a new anti-cancer strategy¹⁴. However, progress in achieving this goal has been hampered by the complexity and potential redundancy of the mediators. Our findings indicate that, in spite of such complexity, blocking the action of a single cytokine, Bv8, may have a significant affect on growth of multiple tumour types. Thus, Bv8 or its receptors may represent therapeutic targets. Additional studies are required to define further the molecular mechanisms of this signalling system and its role in different tumour types and at different stages of tumour progression.

Interestingly, recent studies have shown that administration of G-CSF may induce mouse Vegf expression in neutrophils⁴¹ and accelerates tumour growth^{33,42}. Further studies are warranted to investigate the role of the G-CSF–Bv8 link in tumorigenesis. Also, the finding that Bv8 expression is exquisitely responsive to G-CSF suggests that Bv8 might have a broader pathophysiological role, including involvement in non-tumoural types of neutrophil-mediated angiogenesis⁴³.

METHODS SUMMARY

Expression of Bv8 and its receptors was evaluated by Taqman analysis. To measure Bv8 levels in various tissues and in the serum, we used an ELISA. To test the effects of anti-G-CSF antibodies on Bv8 gene expression induced by the tumour environment, mouse BMMNCs were treated with lysates of HM7 tumours or with control buffer. Tumour lysates had been pre-incubated with a goat anti-G-CSF-neutralizing polyclonal IgG or with control goat IgG at various concentrations. For *in vivo* neutralization of G-CSF, we used the polyclonal IgG or a rat monoclonal antibody. To establish directly the role of Bv8 during tumorigenesis, we used neutralizing anti-Bv8 monoclonal antibodies. The human tumour cell lines Calu6, A673, Jurkat, HPAC and HM7 were injected subcutaneously into the dorsal flank of immunodeficient mice. The mouse lymphoma lines EL4 and TIB42 were injected into immunodeficient or

immunocompetent mice. Animals received anti-Bv8 (monoclonal antibodies 2B9 and 3F1), anti-Ragweed monoclonal antibody and anti-Vegf monoclonal antibody (G6-31 or B20; ref. 32) by intraperitoneal administration. Tumour volumes were calculated every second day⁴⁴. For statistical analysis of differences between groups, a one-way ANOVA followed by a Tukey HSD pairwise analysis was performed using JMP software (SAS Institute Inc.). A P value < 0.05 was considered significant.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 July; accepted 1 October 2007.

- Ferrara, N. & Kerbel, R. S. Angiogenesis as a therapeutic target. *Nature* **438**, 967–974 (2005).
- Carmeliet, P. Angiogenesis in health and disease. *Nature Med.* **9**, 653–660 (2003).
- Ferrara, N., Gerber, H. P. & LeCouter, J. The biology of VEGF and its receptors. *Nature Med.* **9**, 669–676 (2003).
- Folkman, J. & Klagsbrun, M. Angiogenic factors. *Science* **235**, 442–447 (1987).
- Dong, J. *et al.* VEGF-null cells require PDGFR α signaling-mediated stromal fibroblast recruitment for tumorigenesis. *EMBO J.* **23**, 2800–2810 (2004).
- Orimo, A. *et al.* Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. *Cell* **121**, 335–348 (2005).
- Coussens, L. M. & Werb, Z. Inflammation and cancer. *Nature* **420**, 860–867 (2002).
- Raffi, S., Lyden, D., Benezra, R., Hattori, K. & Heissig, B. Vascular and haematopoietic stem cells: novel targets for anti-angiogenesis therapy? *Nature Rev. Cancer* **2**, 826–835 (2002).
- Nolan, D. J. *et al.* Bone marrow-derived endothelial progenitor cells are a major determinant of nascent tumor neovascularization. *Genes Dev.* **21**, 1546–1558 (2007).
- De Palma, M., Veneri, M. A., Roca, C. & Naldini, L. Targeting exogenous genes to tumor angiogenesis by transplantation of genetically modified hematopoietic stem cells. *Nature Med.* **9**, 789–795 (2003).
- Yang, L. *et al.* Expansion of myeloid immune suppressor Gr⁺CD11b⁺ cells in tumor-bearing host directly promotes tumor angiogenesis. *Cancer Cell* **6**, 409–421 (2004).
- De Palma, M. *et al.* Tie2 identifies a hematopoietic lineage of proangiogenic monocytes required for tumor vessel formation and a mesenchymal population of pericyte progenitors. *Cancer Cell* **8**, 211–226 (2005).
- Grunewald, M. *et al.* VEGF-induced adult neovascularization: recruitment, retention, and role of accessory cells. *Cell* **124**, 175–189 (2006).
- Shojaei, F. *et al.* Tumor refractoriness to anti-VEGF treatment is mediated by CD11b⁺Gr1⁺ myeloid cells. *Nature Biotechnol.* **25**, 911–920 (2007).
- Nozawa, H., Chiu, C. & Hanahan, D. Infiltrating neutrophils mediate the initial angiogenic switch in a mouse model of multistage carcinogenesis. *Proc. Natl Acad. Sci. USA* **103**, 12493–12498 (2006).
- van Kempen, L. C., de Visser, K. E. & Coussens, L. M. Inflammation, proteases and cancer. *Eur. J. Cancer* **42**, 728–734 (2006).
- Bergers, G. *et al.* Matrix metalloproteinase-9 triggers the angiogenic switch during carcinogenesis. *Nature Cell Biol.* **2**, 737–744 (2000).
- Li, M., Bullock, C. M., Knauer, D. J., Ehler, F. J. & Zhou, Q. Y. Identification of two prokineticin cDNAs: recombinant proteins potentially contract gastrointestinal smooth muscle. *Mol. Pharmacol.* **59**, 692–698 (2001).
- LeCouter, J. *et al.* Identification of an angiogenic mitogen selective for endocrine gland endothelium. *Nature* **412**, 877–884 (2001).
- LeCouter, J. *et al.* The endocrine-gland-derived VEGF homologue Bv8 promotes angiogenesis in the testis: localization of Bv8 receptors to endothelial cells. *Proc. Natl Acad. Sci. USA* **100**, 2685–2690 (2003).
- Mollay, C. *et al.* Bv8, a small protein from frog skin and its homologue from snake venom induce hyperalgesia in rats. *Eur. J. Pharmacol.* **374**, 189–196 (1999).
- Masuda, Y. *et al.* Isolation and identification of EG-VEGF/prokineticins as cognate ligands for two orphan G-protein-coupled receptors. *Biochem. Biophys. Res. Commun.* **293**, 396–402 (2002).
- Lin, D. C. *et al.* Identification and molecular characterization of two closely related G protein-coupled receptors activated by prokineticins/endocrine gland vascular endothelial growth factor. *J. Biol. Chem.* **277**, 19276–19280 (2002).
- LeCouter, J., Lin, R. & Ferrara, N. Endocrine gland-derived VEGF and the emerging hypothesis of organ-specific regulation of angiogenesis. *Nature Med.* **8**, 913–917 (2002).
- LeCouter, J., Zlot, C., Tejada, M., Peale, F. & Ferrara, N. Bv8 and endocrine gland-derived vascular endothelial growth factor stimulate hematopoiesis and hematopoietic cell mobilization. *Proc. Natl Acad. Sci. USA* **101**, 16813–16818 (2004).
- Dorsch, M. *et al.* PK1/EG-VEGF induces monocyte differentiation and activation. *J. Leukoc. Biol.* **78**, 426–434 (2005).
- Dahl, R. *et al.* Regulation of macrophage and neutrophil cell fates by the PU.1:C/EBP α ratio and granulocyte colony-stimulating factor. *Nature Immunol.* **4**, 1029–1036 (2003).

28. Lagasse, E. & Weissman, I. L. Flow cytometric identification of murine neutrophils and monocytes. *J. Immunol. Methods* **197**, 139–150 (1996).
29. Metcalf, D. The molecular control of cell division, differentiation commitment and maturation in haemopoietic cells. *Nature* **339**, 27–30 (1989).
30. Christopher, M. J. & Link, D. C. Regulation of neutrophil homeostasis. *Curr. Opin. Hematol.* **14**, 3–8 (2007).
31. Mueller, M. M. & Fusenig, N. E. Tumor–stroma interactions directing phenotype and progression of epithelial skin tumor cells. *Differentiation* **70**, 486–497 (2002).
32. Liang, W. C. *et al.* Cross-species VEGF-blocking antibodies completely inhibit the growth of human tumor xenografts and measure the contribution of stromal VEGF. *J. Biol. Chem.* **281**, 951–961 (2006).
33. Okazaki, T. *et al.* Granulocyte colony-stimulating factor promotes tumor angiogenesis via increasing circulating endothelial progenitor cells and Gr1+CD11b+ cells in cancer animal models. *Int. Immunol.* **18**, 1–9 (2006).
34. Garcia-Sanz, A., Rodriguez-Barbero, A., Bentley, M. D., Ritman, E. L. & Romero, J. C. Three-dimensional microcomputed tomography of renal vasculature in rats. *Hypertension* **31**, 440–444 (1998).
35. Maehara, N. Experimental microcomputed tomography study of the 3D microangioarchitecture of tumors. *Eur. Radiol.* **13**, 1559–1565 (2003).
36. Kwon, H. M. *et al.* Enhanced coronary vasa vasorum neovascularization in experimental hypercholesterolemia. *J. Clin. Invest.* **101**, 1551–1556 (1998).
37. Neben, S., Marcus, K. & Mauch, P. Mobilization of hematopoietic stem and progenitor cell subpopulations from the marrow to the blood of mice following cyclophosphamide and/or granulocyte colony-stimulating factor. *Blood* **81**, 1960–1967 (1993).
38. Kavgaci, H., Ozdemir, F., Aydin, F., Yavuz, A. & Yavuz, M. Endogenous granulocyte colony-stimulating factor (G-CSF) levels in chemotherapy-induced neutropenia and in neutropenia related with primary diseases. *J. Exp. Clin. Cancer Res.* **21**, 475–479 (2002).
39. Cheng, M. Y. *et al.* Prokineticin 2 transmits the behavioural circadian rhythm of the suprachiasmatic nucleus. *Nature* **417**, 405–410 (2002).
40. Matsumoto, S. *et al.* Abnormal development of the olfactory bulb and reproductive system in mice lacking prokineticin receptor PKR2. *Proc. Natl Acad. Sci. USA* **103**, 4140–4145 (2006).
41. Ohki, Y. *et al.* Granulocyte colony-stimulating factor promotes neovascularization by releasing vascular endothelial growth factor from neutrophils. *FASEB J.* **19**, 2005–2007 (2005).
42. Hirbe, A. C. *et al.* Granulocyte colony-stimulating factor enhances bone tumor growth in mice in an osteoclast-dependent manner. *Blood* **109**, 3424–3431 (2007).
43. Eyles, J. L., Roberts, A. W., Metcalf, D. & Wicks, I. P. Granulocyte colony-stimulating factor and neutrophils—forgotten mediators of inflammatory disease. *Nature Clin. Pract. Rheumatol.* **2**, 500–510 (2006).
44. Tomayko, M. M. & Reynolds, C. P. Determination of subcutaneous tumor size in athymic (nude) mice. *Cancer Chemother. Pharmacol.* **24**, 148–154 (1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Flow Cytometry Laboratory and the Animal Care Facility for their help. We express our appreciation to G. Fuh, A. Chuntarapai, K. Schroeder and the Antibody Technology group. We acknowledge S. Beddha for immunohistochemistry. We also thank J. LeCouter and M. Singh for reading the manuscript.

Author Contributions F.S., X.W. and C.Z. contributed equally to this work. F.S. and N.F. wrote the manuscript. F.S., X.W., C.Z., L.Y., X.-H.L., J.Y., D.B., C.B., F.V.P., N.v.B., C.H., J.R., M.T., R.A.D.C., Y.G.M. and N.F. performed experiments, designed research and analysed data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.F. (nf@gene.com).

METHODS

Gene expression analysis by Taqman. RNA from tissues or cells was prepared using the RNeasy Mini Kit (Qiagen). For real time PCR (Taqman) analysis, 50 ng total RNA was used for each reaction. For detection of mouse or human *Bv8* and *EG-VEGFR/PKR1*, mouse and human testis RNAs (BD Biosciences), respectively, were used as controls. For mouse or human *EG-VEGFR/PKR2*, mouse and human hypothalamus or whole brain (BD Biosciences), respectively, were used as controls. Reactions were run on the 9,600 emulsion mode of the 7,500 real time PCR system (Applied Biosystems), and the absolute quantification with standard curve was used with sequence detection system (SDS) software. The expression level of each gene was further quantified against the house-keeping gene *Rpl19* or *Gapdh* in the same treatment. To confirm the expression of *VEGFR-1*, *VEGFR-2*, prokineticin receptor 1 (*PKR1*; *EG-VEGFR/PKR1*) and *PKR2* (*EG-VEGFR/PKR2*) in the tumour-associated endothelial cells, standard PCR with reverse transcription (RT-PCR) was conducted using the Titan One Tube RT-PCR system (Roche), and the end products were checked on 2% agarose gel (Invitrogen) for the correct size.

The sequence of Taqman primers were as follows: mouse *Bv8* forward, GCATGACAGGAGTCATCATTTT; reverse, AAATGGCAGGATATCAGGAAA; probe, AAACITTTATTGTAAACCAAGGTCTAATGTAAATGGA. Human *Bv8* forward, ATGGCAGGAAGCTAGGA; reverse, GCAGAGCTGAAGTCCTCTTGA; probe, TGCTGCTGGACCTTCTTAAACCT. Mouse *Bv8 R1* forward, CAGCGCATGAAGACTTG; reverse, GTCATCTTCGGTTTCTGAGT; probe, TCCAGGCAGCACCCCTGATG. Mouse *Bv8 R2* forward, GAACTCCACGTCGAGCGCA; reverse, GGGTCCCAGTTGATGATGC; probe, CTCCTGATACACACAGCCACCTG. Human *Bv8 R1* forward, CTGGAAGGCTTCTTACAATGG; reverse, GGCATCCCAATTGTCTTGA; probe, TCCAGGTCTGCACTGGACTTACCG. Human *Bv8 R2* forward, TCACCATCGTTCGTGACTTC; reverse, AGAAGGCAGTGAGGTAGTGCTT; probe, TCCTTCACGAACACAGTGGGAA. Mouse *Rpl19* forward, AGGTCAAAGGGAATGTGTTCAAA; reverse, CCTTGTCTGCCTTCAGCTTGT; probe, ACAAGCGCATCCTCATGGAGCACATC. Human *RPL19* forward, CGCAAGCGCCGTGAA; reverse, GGTCTCTTCTCCTTGGATAAAGTC; probe, CCAGGCCAAGAAGGAGGAGATCATCA. Human *GAPDH* forward, TGGGCTACACTGAGCACCAG; reverse, CAGCGTCAAAGGTGGAGGAG; probe, TGGTCTCTCTGACTTC-AACAGCGACAC. Mouse *Gapdh* forward, ATGTTCCAGTATGACTCCAC-TACAG; reverse, GAAGACACAGTAGACTCCACGACA; probe, AAGCCCAT-CACCATCTTCCAGGAGCGAGA.

Flow cytometry. BMMNCs, peripheral blood mononuclear cells (PBMCs) and tumour cells were harvested from mice implanted with several tumour types. Red blood cells were lysed using ACK (Cambrax) lysis buffer, followed by staining with rat anti-mouse CD11b (Mylentni Biotech) conjugated to allophycocyanin (APC) and rat anti-mouse Gr1 (BD Biosciences) conjugated to phycoerythrin (PE). To exclude dead cells, 7AAD (aminoactinomycin D; BD Biosciences) was added to all samples before data acquisition in the FACS instrument (BD Biosciences).

Migration assay. BMMNCs were isolated from naive Beige nude mice and the CD11b⁺Gr1⁺ population was sorted using CD11b microbeads (Miltenyi Biotech) according to protocols provided by the manufacturer. An aliquot of the sorted cells was stained with anti-CD11b-APC and anti-Gr1-PE to ensure the purity (more than 99%) of CD11b⁺Gr1⁺ cells. For the migration assay, 2.0×10^5 cells were plated on the top chamber of transwells (Corning Incorporated). The bottom chambers contained 600 μ l of media (IMDM, Gibco BRL) supplemented with BIT (Stem Cell Technologies) containing human BV8, control antibody and mouse recombinant Vegf in separate wells. Cells were incubated in a humidified atmosphere at 37 °C and 5% CO₂ for 9 h, and migration of CD11b⁺Gr1⁺ cells was measured by counting cells in the bottom chamber.

Regulation of Bv8 expression in cultured BMMNCs. Recombinant mouse MCP-1 (also known as Ccl2), MIP-1 α (Ccl3), MIP-1 β (Ccl4), MIP-2 (Ccl2), basic fibroblast growth factor (bFGF), Vegf, GM-CSF, G-CSF, Sdf1 and tumour necrosis factor α (TNF α) were purchased from R&D Systems. Recombinant mouse KC (also known as Cxcl1), interferon α (IFN α), Bv8, IL-4, IL-10, IL-13, transforming growth factor β (TGF- β) were from PeproTech. All cytokines were used at 10 ng ml⁻¹ except for Vegf and Bv8, which were tested at 50 ng ml⁻¹. BMMNCs were flushed from femurs and tibias of Balb/c nude mice with DMEM containing 10% FBS. Cells were centrifuged at 350g for 5 min, and were resuspended in HBSS media containing 0.2% BSA (low endotoxin, Serologicals Corp.) and supplemented with the above cytokines. Two-million cells were incubated in 24-well plates for 4 h at 37 °C in a 5% CO₂ incubator. Total RNA was extracted from each treatment using the RNeasy kit (Qiagen) and *Bv8* expression was assessed by Taqman as described. To investigate *Bv8* expression in the myeloid subset, populations of bone marrow CD11b⁺Gr1⁺ or CD11b⁻Gr1⁻ cells were obtained using FACS.

To test the effects of anti-G-CSF antibodies on *Bv8* gene expression induced by the tumour environment, HM7 tumours were harvested from mice three days after tumour cell inoculation. Tumours were lysed and aliquots were pre-incubated with a goat anti-G-CSF neutralizing polyclonal IgG (AF-414-NA, R&D Systems) or control goat IgG (R&D Systems) at various concentrations for 45 min. Next, naive BMMNCs (isolated from Balb/c nude mice) were treated for 4 h with the HM7 tumour lysates as described above. Expression of *Bv8* in BMMNCs was subsequently evaluated by Taqman, using nine animals from three independent studies.

Collection of condition medium from tumour cells. A673, HM7, HPAC and Calu6 cells were grown in a humidified incubator at 37 °C and 5% CO₂. The growth media, supplemented with 10% FBS, were as follows: high-glucose DMEM and 2 mM glutamine for A673 and HM7 cells; high-glucose DMEM plus 2 mM glutamine, 0.1 mM nonessential amino acids and 1.0 mM sodium pyruvate for Calu6 and HPAC cells; and RPMI 1640 plus 2 mM glutamine, 1.0 mM sodium pyruvate and 10 mM HEPES for Jurkat cells. After reaching 90% confluence, the media were changed to 0.5% FBS-containing DMEM:F12 (50:50). The conditioned media were collected after a three-day incubation, and cell viability and total cell number were measured using Vi-Cell XR (Beckman Coulter).

Tumour cell proliferation assay. A673, HM7, HPAC and Calu6 cells were trypsinized and washed in media containing 0.5% FBS before seeding into 96-well black Viewplate (Packard Bioscience Company). Cells were incubated with various amounts of Bv8 (PeproTech) for three days. Cell proliferation was assessed by BrdU incorporation using the Cell Proliferation ELISA kit (Roche).

In vivo G-CSF and anti-G-CSF studies. Eight-week-old Balb/c mice were injected subcutaneously with 10 μ g recombinant human G-CSF (Neupogen, Amgen) daily for eight consecutive days. Control animals were given PBS. At the end of study, bone marrow, whole-blood and spleen samples were taken for analysis. One group of animals was maintained for two days after the last injection of G-CSF. Neutrophils were counted using an automated, high-resolution, flow-cytometry-based haematology analyser (CellDyn 3000). Levels of Bv8 in the serums and bone marrows were measured by ELISA as described.

To determine the role of Bv8 in G-CSF-induced mobilization of CD11b⁺Gr1⁺ cells, Balb/c nude mice received two doses of anti-Bv8 antibodies (5 mg kg⁻¹ plus 5 mg kg⁻¹) 12 h apart, followed by mouse G-CSF (R&D Systems; 2 or 10 μ g per mouse) 4 h after the second treatment. As a positive control, we used a rat anti-mouse G-CSF monoclonal antibody (Mab414, R&D Systems; 10 or 30 μ g per mouse) given at the same interval as anti-Bv8, followed by mouse G-CSF. After 6 h, mice were bled and the frequency of CD11b⁺Gr1⁺ cells was determined as described.

To investigate the role of G-CSF in regulating Bv8 expression in the absence of tumour, Balb/c nude mice were given daily intraperitoneal injections of control rat IgG (Genentech Inc.) or rat anti-G-CSF monoclonal antibody (R&D Systems; 10 μ g per mouse) for eight consecutive days. Animals were killed and total proteins were extracted from BMMNCs as described, and Bv8 levels were measured by ELISA. To assess the significance of G-CSF in regulating Bv8 expression in tumours, Balb/c nude mice were pretreated with 10 μ g rat anti-G-CSF monoclonal antibody or rat IgG as described above, followed by implantation of HM7 cells (5×10^6 per mouse) after 12 h. Controls were implanted with empty Matrigel. Animals then received daily administration of antibodies for two days. Mice were killed 48 h after Matrigel or tumour implantation, and Bv8 levels in BMMNCs were measured as described.

Generation and screening of anti-Bv8 neutralizing monoclonal antibodies.

Mouse monoclonal antibodies directed against recombinant human BV8 protein were screened using two independent assays. The first assay was based on the ability of BV8 protein to induce proliferation of bovine adrenal cortex-derived endothelial cells, as described²⁰. The second assay relied on the ability of BV8 to induce a signalling cascade in Chinese hamster ovary (CHO) cells stably expressing each of the BV8 receptors. In brief, CHO cells stably expressing the β -lactamase gene under the NFAT promoter (Invitrogen) were grown in DMEM supplemented with 10% fetal bovine serum. Human EG-VEGFR/PKR1 or EG-VEGFR/PKR2 complementary DNA^{22,23} in pMSCV vector with a hygromycin-resistant gene were transduced. Cells expressing the transgene were selected in 500 μ g ml⁻¹ hygromycin for two weeks. Cells responding to the treatment were subsequently isolated by FACS sorting for their ability to cleave the FRET-based fluorescent substrate CCF4 after 16 h stimulation with human BV8, as suggested by the manufacturer. Neutralizing antibodies were identified for their ability to block Bv8-induced β -lactamase expression. CHO-NFAT β -lactamase R1 or R2 were stimulated by human BV8 at 100–200 ng ml⁻¹ for 16 h in the presence or absence of purified mouse monoclonal anti-Bv8 antibodies at various concentrations. After stimulation, cells were incubated with CCF4 for 1 h, and the levels of fluorescence were measured in a 96-well plate reader Envision (Perkin Elmer).

To establish directly the role of Bv8 during tumorigenesis, we used neutralizing anti-Bv8 monoclonal antibodies. Mouse monoclonal antibodies 3F1 and 2B9, which cross-react with mouse and human Bv8, were used. These monoclonal antibodies were selected on the basis of their ability to inhibit Bv8-stimulated adrenal cortex endothelial cell proliferation²⁰. Monoclonal antibody 2B9 maximally inhibited ~70% of the mitogenic effect of human or mouse Bv8 proteins, whereas monoclonal antibody 3F1 inhibited as much as 50%. However, the combination of the two monoclonal antibodies, each at the concentration of 5–10 $\mu\text{g ml}^{-1}$, completely blocked the mitogenic effects elicited by 100 ng ml^{-1} human or mouse Bv8. The antibodies, tested alone or in combination, had no effect on endothelial cell proliferation under basal conditions or after stimulation with the structurally unrelated mouse and human VEGF or the related EG-VEGF. Also, neither the anti-Bv8 monoclonal antibodies nor Bv8 itself had any detectable effects on the proliferation of the tumour cell lines tested in this study, over a wide range of concentrations (data not shown).

To determine the most effective therapeutic regimen *in vivo*, in initial experiments we performed dose–response studies with monoclonal antibodies 3F1 and 2B9, individually and in combination, in the A673 model. As predicted by the *in vitro* data, a combination of the two monoclonal antibodies was more effective than a single monoclonal antibody. Administration of 5 mg kg^{-1} of each monoclonal antibody twice weekly achieved a maximal inhibitory effect on tumour growth. Therefore, this regimen was used in the proof-of-concept experiments shown in this manuscript. Additional anti-Bv8 antibodies have been generated and characterized.

***In vivo* tumour studies.** The human tumour cell lines Calu6, A673, Jurkat, HPAC and HM7 as well as the mouse lymphoma lines EL4 and TIB42 were grown in DMEM:F12 (50:50) media supplemented with 10% (v/v) FBS, 1% (v/v) penicillin and streptomycin, 2 mM L-glutamine and 1 $\mu\text{g ml}^{-1}$ fungizone (Invitrogen). Tumour cell lines were grown by incubating at 37 °C and 5% CO₂ in a humidified atmosphere. For mouse xenograft experiments, tumour cells were resuspended at a concentration of 1×10^8 cells per ml Matrigel (growth-factor-reduced; BD Pharmingen) and injected (100 μl) subcutaneously into the dorsal flank of either Balb/c nude or Beige nude mice (Harlan Sprague Dawley). Anti-Bv8 treatment was started (5 mg kg^{-1} of each 2B9 and 3F1 monoclonal antibodies) 24 or 48 h after tumour cell inoculation by the intraperitoneal route of administration. Anti-ragweed monoclonal antibody (10 mg kg^{-1}) and anti-Vegf monoclonal antibody G6.31 or B20 (5 mg kg^{-1})³² served as controls. Monoclonal antibodies were administered to tumour-bearing mice twice weekly and tumour volumes were calculated every other day using the ellipsoid volume formulas ($0.5 \times L \times W^2$, where L is length and W is width)⁴⁴. For statistical analysis, a one-way ANOVA followed by a Tukey HSD pairwise analysis was performed using JMP software (SAS Institute Inc.), and P values ≤ 0.05 were considered significant.

Histological analysis and immunohistochemistry. Tumours were fixed in neutral-buffered formalin for 24 h before paraffin embedding. Haematoxylin and eosin staining and immunohistochemistry were performed as described previously²⁵. Briefly, immunohistochemical staining with rat anti-mouse PLVAP monoclonal antibody MECA-32 (BD Pharmingen) was performed using target antigen retrieval solution (DAKO) at 99 °C followed by incubation at room temperature (22 °C) for 20 min. Next, sections were stained with a biotinylated secondary antibody (Vector Laboratories) for 30 min at room temperature followed by incubation with Vectastain ABC Elite reagents. Sections were then incubated with peroxidase substrate solution (metal-enhanced DAB, Pierce Chemical), until the desired intensity was developed. Finally, sections were lightly counterstained with haematoxylin, dehydrated, and were coverslipped.

For anti-Gr1 immunohistochemistry, acetone-fixed 5- μm -thick frozen tumour sections were blocked with 10% goat serum for 30 min. Sections were incubated with rat monoclonal anti-Gr1 antibody (clone RB6-8C5, BD Pharmingen) at 6 $\mu\text{g ml}^{-1}$ for 1 h at room temperature, followed by Cy2-conjugated goat anti-rat (Jackson Immuno Research) at 5 $\mu\text{g ml}^{-1}$ for 30 min at room temperature. Sections were counterstained with DAPI and coverslipped with ProLong Gold mounting media (Invitrogen).

Construction of adenovirus vectors. Adenovirus vectors encoding LacZ and mouse Vegf₁₆₄ were described previously¹⁹. Adenoviral mouse Bv8 was generated using the AdEasy XL adenoviral vector system (Stratagene). The cDNA of mouse Bv8 with a 6 \times His tag at its carboxy terminus was cloned between the XhoI and HindIII sites of the pShuttle-CMV vector. The resultant pShuttle-CMV-Bv8 plasmid was recombined with pAdEasy-1 in BJ5183-AD-1, an electroporation-competent strain pre-transformed with the adenoviral backbone. The recombinant adenoviral Bv8 plasmid was then transfected into AD-293 cells for packaging viral particles. Adenovirus stocks were purified by CsCl gradient and were titrated using Adeno-X rapid titer kit (Clontech).

Isolation and characterization of tumour-associated endothelial cells. Tumour-associated endothelial cells (TAECs) were isolated using a magnetic

bead sorting system (Miltenyi Biotech), essentially as described previously⁴⁵. In brief, TIB42 mouse lymphoma cells were injected subcutaneously into the dorsal lateral flanks of female Beige nude mice. Animals were killed when tumours reached ~1,000 mm^3 in diameter. The harvested tumours were excised, minced and then digested with collagenase II (Worthington Biochemical Corporation). Tumour cell suspensions were then filtered using cell strainers (100 μm and 40 μm pore size; BD Falcon) to exclude debris and undigested tumours. The remaining cells were stained with a FITC-conjugated anti-CD31 antibody (BD Biosciences) and were sorted in a FACSARIA instrument (BD Biosciences). The CD31⁺ sorted cells were seeded in gelatin-coated plates in the presence of EGM-2MV media (Cambrex), and after 24 h the non-adherent cells were removed by washing the plates several times with PBS.

Total RNA was extracted from cultured cells using the RNeasy minikit (Qiagen). For PCR analysis, 80 ng of the total RNA was used for each reaction and the cDNA was amplified for 28 cycles. The identity of TIB42 TAECs was confirmed by a variety of criteria (See Supplementary Fig. 8). Additionally, cultured TIB42 TAECs expressed both EG-VEGFR/PKR1 and EG-VEGFR/PKR2 (data not shown). To assess MAP kinase activation, TIB42 TAECs were starved for 6 h in basal media supplemented with 0.5% BSA. Cells were then stimulated with human recombinant BV8 (200 ng ml^{-1} ; Peprotech), complete media, human VEGF (100 ng ml^{-1} ; Peprotech) or BSA (0.5%). Cell extracts were collected at the indicated time points; that is, 5, 10 and 20 min after stimulation. Western blot analysis of the extracts from TIB42 TAECs was performed using PhosphoPlus p44/42 MAPK kit (Cell Signaling). To evaluate the consistency and reproducibility of the result, each condition was tested in duplicate, and the experiments were performed three times.

For *in vitro* tube formation, TIB42 TAECs (passages 6–8) were starved for 5 h in serum-free medium. Cells were then collected and resuspended in serum-free medium supplemented with 5% BSA and treated with human VEGF (100 ng ml^{-1}), BV8 (200 ng ml^{-1}) or sham control (control). Treated cells (5×10^5) were seeded in Matrigel (BD Biosciences) pre-coated wells, and tube formation was evaluated after 36 h incubation in a humidified incubator at 37 °C and 5% CO₂.

Micro-CT angiography. To assess the tumour vasculature, we used X-ray micro-CT^{34–36}. Micro-CT provides an overall analysis of tumour vasculature in the entire tumour and thus may overcome some limitations inherent in some other approaches such as immunohistochemistry.

HM7-tumour-bearing animals received a 50 μl intraperitoneal injection of heparin 10 min before being killed by inhalation of carbon dioxide. The thoracic cavity was opened, an incision was made in the apex of the heart, and a polyethylene cannula (inner diameter, 0.58 mm; outer diameter, 0.96 mm) was passed through the left ventricle and secured in the ascending aorta with a 5-0 silk suture. A 17 ml solution of 0.1 mM sodium nitroprusside in 0.9% saline was perfused at a rate of 6 ml min^{-1} to provide a state of maximum vasodilatation and to remove blood. MICROFIL (Carver), a commercially available lead chromate latex, was prepared as recommended by the manufacturer. Mice were then perfused with 17 ml of MICROFIL at a rate of 2 ml min^{-1} . The infused latex mixture was allowed to polymerize at room temperature for sixty minutes before dissection of tissues of interest. Dissected tumours were immersed in 10% neutral buffered formalin.

The tumours were then imaged with a $\mu\text{CT}40$ (SCANCO Medical) X-ray micro-CT system. The tumours were imaged with soybean oil as the background media. The micro-CT images were generated by operating the X-ray tube at an energy level of 45 kV, a current of 177 μA and an integration time of 300 ms. Axial images were obtained at an isotropic resolution of 16 μm .

The vascular network and tumour were extracted by a series of image-processing steps. An intensity threshold of 1,195 Hounsfield units and morphological filtering (erosion and dilation) were applied to the volumetric micro-CT image data to extract the VV. The TV was extracted from the background in similar fashion with an intensity threshold of -8 Hounsfield units. Vessel density (VV/TV) was determined from the ratio of VV to TV. The vascular and tumour intensity thresholds were determined by visual inspection of the segmentation results from a subset of samples. Computations were performed by an in-house image analysis algorithm written in C++ and Python that used the AVW image-processing software library (AnalyseDirect Inc.). Three-dimensional surface renderings were created from the micro-CT data with the use of Analyse 6.0 (AnalyseDirect Inc.), an image analysis software package. Statistical analysis was performed with JMP statistical software package (SAS Institute Inc.). Group comparisons for micro-CT metrics (VV, TV, VV/TV) were evaluated with Dunnett's test for multiple comparisons. P values less than 0.05 were considered significant.

Partial purification of Bv8 protein. Balb/c mice ($n = 20$) were injected subcutaneously with human G-CSF (10 $\mu\text{g day}^{-1}$, Neupogen, Amgen) and were injected intraperitoneally with mouse GM-CSF (0.5 $\mu\text{g day}^{-1}$, Peprotech) daily

for four days to expand the CD11b⁺Gr1⁺ population. On day five, the bone marrow cells were isolated and the cell pellet was resuspended in 2 ml of 0.5% Triton X-100. The cell lysate was then forced through a 25-gauge needle four times, and the salt concentration was adjusted to 50 mM NaCl. The crude extract was applied to a heparin–Sepharose column (Hi-Trap, 1 ml) pre-equilibrated with 20 mM Tris pH 7.2, 50 mM NaCl. The column was eluted using a two-step linear gradient: 50 mM to 1 M NaCl, and then 1 M to 2 M NaCl, in 20 mM Tris, pH 7.2. The flow rate was 1 ml min⁻¹. Absorbance was monitored at 280 nm. Fractions of 1 ml were collected and assayed for mouse Bv8 using ELISA and western blot analysis. For western blot analysis, 100 µl of fractions were concentrated 4-fold using Microcon YM-3 spin column (Millipore), and then loaded on a 4–20% SDS–PAGE (Invitrogen). The blot was stained overnight (8–12 hours) using a combination of three hamster antibodies against mouse Bv8 (2D3, 3B8 and 4E10) at a total concentration of 10 µg ml⁻¹ in blocking buffer (PBST, 0.1% Tween 20 in PBS and 5% skimmed milk). Following three washes, the blot was incubated with a horseradish-peroxidase-conjugated goat anti-hamster IgG (Jackson ImmunoResearch Laboratories), and was then developed using the enhanced chemiluminescence plus western blotting detection system (GE Healthcare).

Mouse Bv8 ELISA. MaxiSorp 96-well microwell plates (Nalge Nunc International) were coated with 1.0 µg ml⁻¹ 3F1 monoclonal antibody in 50 mM carbonate buffer, pH 9.6, at 4 °C overnight. Plates were washed with PBS containing 0.05% polysorbate 20 and blocked with 0.5% BSA, 15 p.p.m. (parts per million) Proclin 300 (Supelco) in PBS at room temperature for 1 h. After plates were washed, mouse Bv8 standards (0.0195–2.50 ng ml⁻¹ in 2-fold serial dilution, Genentech) and samples (minimum 1:10 dilution) in PBS containing 0.5% BSA, 0.05% polysorbate 20, 15 p.p.m. proclin 300 (Supelco) and 0.35 M NaCl (sample buffer) were serially diluted and were added to each well. The plates were incubated for 2 h at room temperature followed by a washing step to remove the unbound antibody. The bound antibody was detected by adding the secondary antibody, a biotinylated hamster anti-mouse Bv8 antibody clone 4E10 (Genentech Inc.) followed by adding streptavidin-HRP (GE Healthcare) and 3,3',5,5'-tetramethyl benzidine (Kirkegaard & Perry Laboratories) as the substrate. The reaction was stopped by adding 1 M phosphoric acid. Absorbance was read at 450 nm on a Titertek stacker reader (ICN). The titration curves of the standards were calculated using a four-parameter regression curve-fitting program (Genentech Inc.). Mouse Bv8 concentration was calculated by extrapolating the optical density values of samples to the data range in the standard curve. The Bv8 ELISA assay was capable of measuring up to 10% mouse serum and 10% lysis buffer, and had the sensitivity to detect Bv8 as low as 0.195 ng ml⁻¹ in serum and tissue lysate samples. The ELISA assay was specifically developed and optimized for Bv8 measurement because anti-VEGF monoclonal antibody, human G-CSF and human VEGF up to 30 µg ml⁻¹, or human EG-VEGF up to 5 µg ml⁻¹, only gave a background signal. The ELISA assay could also detect human BV8, but with less than 26% sensitivity. The presence of human G-CSF and anti-G-CSF (up to 25 µg ml⁻¹) did not significantly affect detection of 0.5 ng ml⁻¹ BV8 in the buffer. The higher concentration of each of the anti-Bv8 monoclonal antibodies (3F1 at concentration greater than 14 ng ml⁻¹ or 2B9 at concentration greater than 124 ng ml⁻¹) reduced the recovery of mouse Bv8 in the buffer. Samples containing anti-Bv8 antibodies were first diluted ten times before the ELISA assay to report most of the free/unbound Bv8.

45. Hida, K. *et al.* Tumor-associated endothelial cells with cytogenetic abnormalities. *Cancer Res.* **64**, 8249–8255 (2004).

ARTICLES

The molecular architecture of cadherins in native epidermal desmosomes

Ashraf Al-Amoudi¹, Daniel Castaño Díez¹, Matthew J. Betts¹ & Achilleas S. Frangakis¹

Desmosomes are cadherin-based adhesive intercellular junctions, which are present in tissues such as heart and skin. Despite considerable efforts, the molecular interfaces that mediate adhesion remain obscure. Here we apply cryo-electron tomography of vitreous sections from human epidermis to visualize the three-dimensional molecular architecture of desmosomal cadherins at close-to-native conditions. The three-dimensional reconstructions show a regular array of densities at ~ 70 Å intervals along the midline, with a curved shape resembling the X-ray structure of C-cadherin, a representative 'classical' cadherin. Model-independent three-dimensional image processing of extracted sub-tomograms reveals the cadherin organization. After fitting the C-cadherin atomic structure into the averaged sub-tomograms, we see a periodic arrangement of a *trans* W-like and a *cis* V-like interaction corresponding to molecules from opposing membranes and the same cell membrane, respectively. The resulting model of cadherin organization explains existing two-dimensional data and yields insights into a possible mechanism of cadherin-based cell adhesion.

Desmosomes are cadherin-mediated intercellular junctions that reinforce hard tissues and encode instructions to drive tissue morphogenesis and to regulate tissue homeostasis¹. Like classical cadherins, desmosomal cadherins (desmocollins and desmogleins) consist of an extracellular domain containing calcium-binding sites between five tandemly repeated domains, a transmembrane domain and a cytoplasmic tail. The extracellular domains of the desmosomal cadherins are homologous to those of the classical cadherins². The molecular interfaces mediating adhesion and specificity reside in the amino-terminal extracellular domain of the cadherins³. Their importance in maintaining tissue integrity is illustrated in various autoimmune and genetic diseases⁴. However, despite numerous structural and mutational studies aiming to identify these interfaces, the arrangement of cadherin molecules underlying adhesion remains controversial.

Here we use cryo-electron tomography⁵ of vitreous sections to visualize the three-dimensional (3D) architecture of desmosomes in human epidermal biopsies. Previous two-dimensional (2D) studies using cryo-electron microscopy of vitreous sections (CEMOVIS) revealed a quasi-periodic organization of cadherin molecules, which seemed straight and densely packed^{6,7}. This is in contrast to the organization obtained with conventional preparation techniques, which showed a stochastic and clustered arrangement of cadherin molecules in the desmosome⁸. The scrutiny of vitrified samples is currently the only way of investigating the macromolecular organization of eukaryotic cells or tissue under close-to-native conditions with molecular resolution. Several technical limitations have been overcome over the past few years^{9,10}, yielding micrographs that faithfully represent the biological material. These micrographs of unstained vitrified samples have a low contrast and a poor signal-to-noise ratio compared with those of stained and dehydrated material prepared with conventional methods¹¹. However, in unstained vitrified samples the intensities represent the underlying true densities of the molecules, allowing for quantitative studies similar to single-particle analysis of cryo-samples¹².

The general structure of the desmosomal region

Desmosomes are readily recognized in 2D micrographs by their lamellar structure and the surrounding complex network of

intermediate filaments (Fig. 1a), but their complete ultrastructure can only be seen from 3D reconstruction. In a reconstruction of a different region of the epidermis, several desmosomes are shown (Fig. 1b, c). They are seen as multi-layers including the two electron-dense cell membranes bisected at the centre of the intercellular space by an electron-dense midline. Furthermore, several organelles can be visualized with unprecedented clarity: the nucleus can be recognized by its nuclear envelope (NE) and nucleoplasm (Nu), which is characterized by the condensed chromatin near the nuclear envelope and distinguished from the interchromatin space by its different texture. The cytoplasm is crowded with several structures: the endoplasmic reticulum (ER) near the nuclear envelope; individual putative ribosomes seen as dense dots; and bundles of tonofilaments, a type of keratin intermediate filament in the cytoskeleton that is characteristic of epithelial tissues (Supplementary Movie 1).

Our analysis of the membrane-to-membrane distance in 24 desmosomes resulted in an average size of 35 nm with a variance of 3 nm. This variation might be natural or might be caused by compression forces during the sectioning process¹³; however, it is worth noting that the intermembrane distance varies significantly within a single desmosome, which can not be accounted for by compression (Supplementary Table 1). Taken together, our measurements are well in the range of the 38.5 nm membrane-to-membrane distance determined by the X-ray structure of the C-ectodomain when cadherin dimers are oriented perpendicular to the membranes¹⁴.

When imaged with a higher magnification and less defocus, rod-like traversing densities can be seen within the extracellular space of the desmosome (Fig. 2; the tilt-series was recorded at a nominal defocus of $-9 \mu\text{m}$, see also Supplementary Table 2). These are ~ 3 nm in diameter and are joined at the junction centre forming the midline. The molecular origin of these densities cannot be unambiguously associated with the available X-ray structures of cadherins, even though the signal-to-noise ratio is sufficient to resolve the bilayer of the cell membranes (Supplementary Fig. 1). However these densities have a clear local periodicity, which can be visualized in Fourier space as well as with autocorrelation functions (Supplementary Fig. 2). Power spectrum measurements reveal that the

¹European Molecular Biology Laboratory, Meyerhofstr. 1, 69117 Heidelberg, Germany.

densities are positioned at ~ 70 Å intervals along the midline and thus correspond to a density of $\sim 17,500$ molecules μm^{-2} . These results are consistent with the molecular packing observed in the X-ray structure of the first domain (EC1) of N-cadherins (neural cadherins)¹⁵. According to this structure, the distance between the two-fold axes of adjacent cadherin dimers protruding from the same cell surface is ~ 75 Å. In some cases, we could recognize W shapes, which closely resemble the X-ray structure of the entire extracellular domain of C-cadherin¹⁴ (Fig. 2).

Sub-tomogram analysis and averaging

To improve the resolution and visualize interactions between the molecules, we performed averaging and subsequent classification of many small regions of the reconstruction, which were extracted around the dense midline and which we call sub-tomograms (Supplementary Fig. 3)¹⁶. We applied this procedure individually to three different reconstructions and verified the results on each one independently. Owing to small variations in the membrane-to-membrane distance of the desmosome, as well as focus differences between the reconstructions, combining the sub-tomograms did not improve the resolution. We chose the tomograms according to a

nominal defocus of -4 μm (Supplementary Table 2) and small alignment errors (less than 0.8 nm), indicating a rigid sample and good alignment. The vitreous sections were particularly thin (~ 50 nm) to allow for high-resolution 3D reconstructions. The resolution was estimated at 3.4 nm by the 0.5 criterion of the Fourier shell correlation, and was close to the first zero-crossing of the contrast transfer function of the microscope (Supplementary Fig. 4).

From each tomographic reconstruction, we automatically extracted ~ 400 sub-tomograms around pixels with the lowest local intensities located close to the dense midline and rotationally aligned them according to their autocorrelation features. We did not translationally align them initially. Because reference bias is a common problem in averaging techniques, we avoided using any template-based methods (for example, cross correlation) for the alignment¹⁷. In the rotationally averaged sub-tomograms a clear periodicity was visible, as expected from observations of the unprocessed data (Fig. 2 and Supplementary Fig. 2). The signal seemed to persist in the midline and faded away when approaching the cell membrane, suggesting that the molecules are stabilized at the midline and have some flexibility farther from it. The midline showed as a dense 70 Å thick layer. This measurement agrees with the size of the adhesive contact region

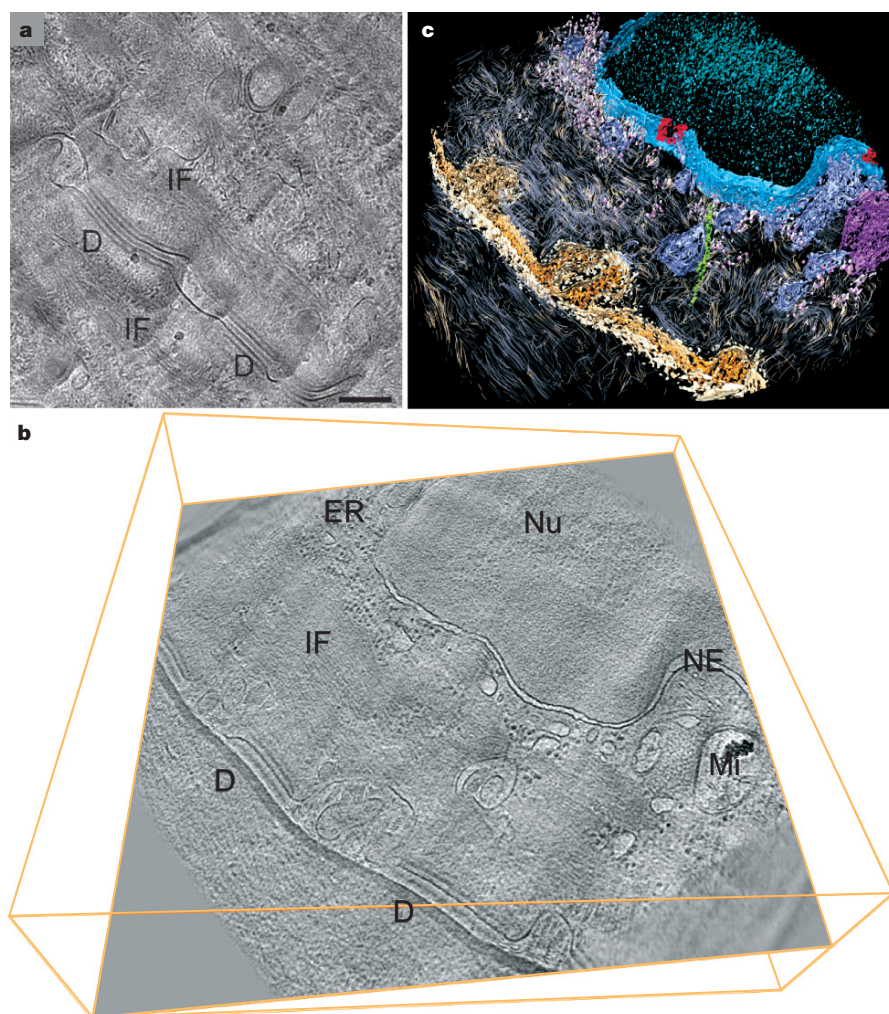


Figure 1 | Visualization of the desmosomal region. **a–c**, Desmosomes (D), nucleus (Nu), nuclear envelope (NE), endoplasmic reticulum (ER), mitochondrion (Mi) and filament networks (IF) in vitreous sections of human epidermis. **a**, A 2D transmission electron micrograph of an 80-nm-thick vitreous section. Scale bar, 200 nm. **b**, A 2.4-nm-thick slice from a 3D reconstruction of a vitreous section of another region of the epidermis (see also Supplementary Movie 1). **c**, Colour-coded isosurface representation of

b after semi-automated segmentation³³. Colours: desmosomal regions, sandy brown; desmosomal plaques, beige; mostly linear features, light steel blue; ER, steel blue; mostly ribosomes, light grey with purple; NE, blue; nuclear pore complexes, red; putative microtubule, green; Mi, purple; and nucleoplasm, dark blue. Protein densities are shown as dark intensities, as recorded from the electron microscope.

observed in the X-ray structure of the entire C-ectodomain¹⁴ and indicates that the adhesive interface in desmosomal junctions is likely to be restricted to the EC1 domain. This is in agreement with the majority of the available data, which indicate that the primary site of adhesive binding in cadherins involves a highly conserved Trp 2 and the specificity-determining residues of EC1 (refs 2 and 18).

Interaction of the extracellular domains

We performed the translational alignment after missing-wedge independent classification techniques identified specific classes¹⁶. The classes revealed orientations in which the curved nature of the cadherin molecules, as well as their organization with respect to their interaction partners, was visible. Two predominant interactions, which were presumably confined to EC1 regions, could be identified (Fig. 3). The first was the *trans* W-like interaction between molecules emanating from opposing cell surfaces. This arrangement closely resembles the *trans* interaction observed in the X-ray structure of C-ectodomain involving the Trp 2 and the hydrophobic pocket¹⁴. The second was the V-like interaction between molecules emanating

from the same cell membrane and thus was considered a possible *cis* interaction because lateral dimerization is considered a fundamental determinant of cell adhesion^{19–22}. In this interaction, the molecules are oriented in parallel such that their concave surfaces face each other, which may be important in maintaining strong intercellular adhesion and specific interactions at the EC1 when the cell moves laterally (for example, under shearing forces). Another interesting feature was the densities observed at the positions close to the inter-domain EC4–EC5 region. Although the limited resolution prevents a firm conclusion, we suggest that this is a *cis* interaction between membrane proximal regions. These were shown to have an important role in adhesion, as suggested both by studies of the effects of the EC5 truncation in desmoglein 1 (ref. 23) and by the generation of inhibitory antibodies associated with the EC4–EC5 region²⁴.

Most of the desmosomal cadherins protrude straight from the cell membrane and are arranged in a regular manner. However, our measurements show that many molecules are flexible, making an angle of up to 20° with the cell membrane (Supplementary Fig. 5). We found that this flexibility is partially associated with the variation in desmosome size. Given this variation, additional flexibility in the angle between the molecular interfaces at the midline is also expected. Although we could not measure this angle precisely on a single molecule basis, X-ray structures from various crystal forms have shown intermolecular angles of the EC1 domains ranging from 54° to 88°^{14,15}. The high flexibility associated with the EC1 domains may

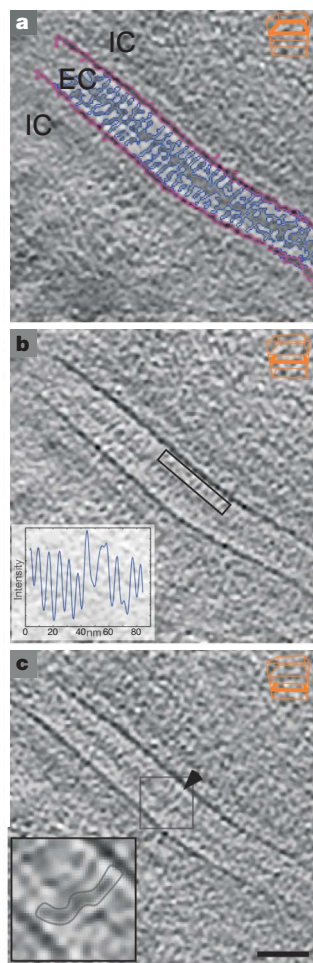


Figure 2 | Tomographic slices of the desmosome. **a–c**, Three 2.4-nm-thick slices with a 7-nm distance from each other. **a**, The cell membranes are highlighted in pink and the densities in between in blue. The densities span the extracellular space (EC) and merge in the middle, forming a continuous layer. In the projection image this layer has the characteristic phenotype of a dense midline of desmosomes between the intracellular spaces (IC). **b**, Inset, a line profile of the boxed region. A clear periodicity is seen in the first half but is disturbed in the latter half. **c**, At specific positions, W-shaped densities can be recognized (indicated by the black arrowhead and magnified in the inset). These span the extracellular space, which indicates a *trans* interaction even though they cannot be unambiguously associated with known X-ray structures. Scale bar, 35 nm.

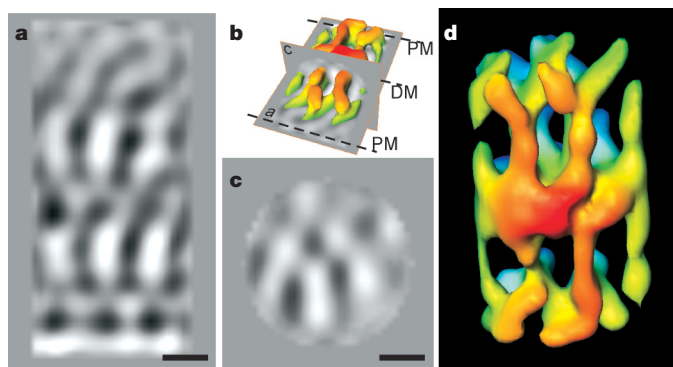


Figure 3 | Visualization of the average of the sub-tomograms. **a**, Coronal slice through the average of the sub-tomograms, which are extracted from the extracellular space only. The plasma membranes are not shown; their position is indicated in **b**, (see also Supplementary Fig. 3 for the exact position of the sub-tomograms with respect to the plasma membranes). The elongated curved densities arrange in a periodic manner and their dimensions compare remarkably well with the W-like conformation observed in the C-cadherin X-ray structure¹⁴, so we consider them to be cadherin molecules. **b**, Cartoon image showing the relative orientations of the slices visualized in **a** and **c** and the isosurface in **d**. The position of the coronal slice visualized in **a**, the axial slice in **c**, and their location with respect to the isosurface shown in **d** are indicated. The dashed lines indicate the location of the dense midline (DM) as well as the position of the cell membranes (PM). The latter are removed from the sub-tomograms and are thus not visible in the average structure or shown in **a**. **c**, Axial slice through the averaged sub-tomograms. Distinct cadherin molecules can be recognized as dense blobs of approximately 3 nm in diameter. **d**, Isosurface visualization of the organization of the cadherin molecules. The threshold is chosen so that the thickness of the central cadherin molecules is approximately 3 nm. The colour coding varies as a function of the depth from red (close) to blue (far). Two cadherin dimers arranged in a *trans* W-like manner are shown in the foreground in orange, with their concave side facing left. One layer deeper, shown in green, four cadherin *trans* W-like dimers have their concave side oriented to the right. The orange and the green molecules emanating from the same cell surface interact to form V-like *cis* dimers (see also Fig. 4). Similarly, two additional W-like cadherin dimers have their concave side oriented to the left (blue) and interact with the green molecules forming V-like *cis* dimers (see also Supplementary Movie 2). Scale bar for **a** and **c**, 7 nm.

counteract shear forces and preserve intercellular adhesion while maintaining specific interactions.

Fitting of atomic structures

To characterize the molecular interfaces involved in the predominant interactions observed in our tomograms, we computationally fitted the entire C-cadherin structure¹⁴ onto our electron microscopy maps by means of cross-correlation techniques. Although desmosomes are heterogeneous junctions consisting of two types of cadherin molecules (desmocollins and desmogleins), the fitting is justified as a result of the high degree of sequence similarity between C-cadherins and desmosomal cadherins². Owing to the curved nature of the molecules, which closely resembles the curvature of C-cadherins, the fitting in the central region of the averaged sub-tomograms is unambiguous despite the missing wedge. Because the identification of the *trans* interface is below our resolution limit, we assumed as an a priori condition for the fit that the side chain of Trp2 is inserted into the hydrophobic pocket as solved by ref. 14. In fact, the arrangement of the densities of the averaged sub-tomograms, which is hampered by the missing wedge, might allow for a fitting such that the Trp 2 is involved in a *cis* rather than a *trans* interaction. However, several studies (that is, site-directed mutagenesis and cell-adhesion assays^{25,26}) have shown that the Trp-2-mediated *trans* interaction is physiologically relevant, so we believe that this is a sensible assumption. Owing to the flexibility of the molecules, the signal fades away from the centre, thereby reducing the quality of the fit. Hence, after the central molecules were unambiguously fitted, their positions were considered as set and all the other molecules were fitted accordingly (Fig. 4, and Supplementary Movie 2). We probed a few other models^{20,27}, but although they might fit in the dimer, they failed when

analysed in the context of the tetramer. This is mainly because of the different arrangement of molecules that is incompatible with our data.

Our results indicate that the molecules interact at the midline, forming building blocks of alternate *cis* and *trans* dimers, and thus resulting in a highly packed regular organization. A high concentration of molecules on the cell surface may be required for strong adhesion in cadherin-based junctions, given the intrinsic weakness of the interaction between individual cadherins suggested by several studies^{19,21,28,29}. According to the recent model of ref. 30, which is based on numerical estimates, high binding affinity would compromise the binding specificity because any two cells expressing cadherins would stick together and disrupt tissue patterning. It is therefore essential that the cadherin concentration at the initial contact of the two cells is kept low so that only a small number of dimers are formed. Once the initial contact is established, more cadherins cluster and adhere to maximize the adhesion and to form a mature desmosome similar to the one we observe here.

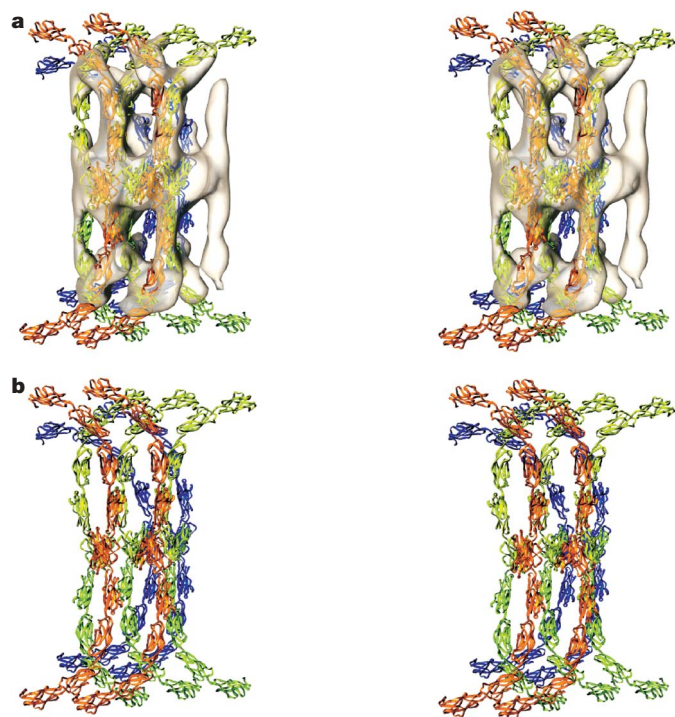


Figure 4 | Stereo view of the C-cadherin X-ray structures¹⁴ fitted to the average of the sub-tomograms (see also Supplementary Movie 2). **a**, The semi-transparent isosurface shown in Fig. 3d with similarly colour-coded cadherin molecules fitted onto the density. The molecules in the front layer are shown in orange, in the middle layer in green and in the rear layer in blue. An unoccupied density is visible on the right side of the isosurface; it represents an additional, incomplete dimer (see also Supplementary Information), and therefore the fitting did not work. **b**, The organization of the cadherin molecules without the isosurface.

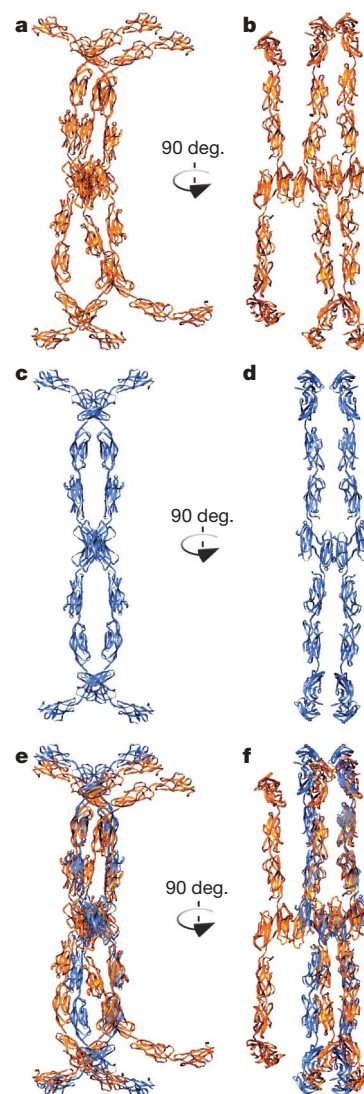


Figure 5 | Comparison of the fitted cadherins with the manually adjusted model of the interfaces of N-cadherin¹⁵. **a**, Fitted cadherins. **b**, Cadherins rotated 90°. The alternating *cis-trans* interactions are visible. **c**, Model. **d**, Model rotated 90°; two *trans* and one *cis* interactions are visible. **e**, Overlap of the fitted cadherin molecules in yellow with the model in blue. The overall good fit is visible, particularly after the 90° rotation (**f**).

Modelling of the desmosomal junction

The organization of *trans* and *cis* interactions at the EC1 shown in our maps has high similarity to the packing in the X-ray structure of the first domain of the N-cadherins¹⁵ (PDB coordinates are 1NCI and 1NCH, respectively). Because this structure lacks the EC2–EC5 domains, we superposed four copies of the C-cadherin ectodomain¹⁴ (PDB 1L3W) on to an N-cadherin tetramer with two *trans* and one *cis* interfaces (from 1NCI and 1NCH, respectively). To approximate the resulting structure to our data, we manually changed the angles of the interfaces by rotations of 30° around the *x* axis of 1NCI and 8° around the *z* axis. This adjustment is justified by the higher similarity to the electron microscopy data, and also by the difference in the angle of the N-cadherin and C-cadherin interfaces¹⁸. This suggests either that the intermolecular angles are flexible or that different cadherins interact at different angles. This arrangement of the tetramer resembled the fitted data very well (Fig. 5, and Supplementary Movie 2). We then expanded these tetramers into a 2D array with the periodicity of the molecules the same as measured in our data. This resulted in a pseudo-atomic model of the desmosome. To verify this model, we created an array of molecules and compared

projection images from different angles with previous electron micrographs of desmosomes recorded with CEMOVIS. Strikingly, all the CEMOVIS images could be reproduced using this rigid model (Fig. 6). Introducing the flexibility found in our data as random variables into the model did not change the projection image enough to prevent the reproduction of the CEMOVIS images. Moreover, projection images along the axis of the dimers closely resemble the images produced by freeze substitution techniques, showing a quasi 2D crystal of the cadherin arrangement³¹. This leads us to believe that our model is indeed correct.

Discussion

In this study we have visualized the 3D organization of desmosomal cadherins under close to native conditions and derived a pseudo-atomic model of their architecture. In contrast to previous tomographic studies with conventional preparation techniques, in which a stochastic arrangement and interaction of cadherins was found⁸, here the cadherins are quasi-periodically arranged and adopt a specific organization with alternating *trans* and *cis* interactions. The differences from the previous tomographic study can be explained by the dehydration procedure, which is shown to cause aggregation³². In the current study, the molecules interact by their tips in a zipper-like arrangement similar to the N-cadherin crystal¹⁵, but in a different manner from the one interpreted when the crystal structure was solved. Despite the quasi-periodicity of the cadherin arrangement, the cadherins retain a significant flexibility without losing their alternating interaction pattern.

Even though the images are static, our results support the hypothesis that desmosomal cadherins on the cell surface are first clustered into small groups interacting through specific residues in the EC1 domains to form *cis* homodimers³⁰. The opposing cell membranes are then brought in close proximity to enable the formation of the *trans* homodimers, which relies on Trp 2 and the hydrophobic pocket together with residues involved in molecular specificity. Once the initial recognition is established, more molecules are brought to the contact zone, thus compacting the junction. This compaction process is regularized by building blocks of alternate *cis* and *trans* dimers so that the strength of cell–cell contact is homogeneous. These processes are repeated to extend the junction and finally form the fully mature desmosome.

METHODS SUMMARY

Skin biopsies were taken from the mid-portion of the forearm of a healthy male who had no history of dermatological disease, and were cryo-fixed with a diameter of 1 mm in the Leica EMPACT2 high-pressure freezer. The frozen samples were transferred to the cryo-chamber of a Leica EM FC6/UC6 ultramicrotome for cryo-sectioning. The sample blocks were prepared using a 20° trimming diamond knife and cryo-sectioned at –140 °C with a nominal section thickness of 50 nm, using a 35° sectioning diamond knife at cutting speed of 1.0 mm s^{–1}. Vitreous sections with good morphology were chosen for data collection. Tilt-series of desmosomal regions of more than 50 cells were collected with a Tecnai Polara microscope, typically covering an angular range of –65° to +65° with 1° or 2° angular increment. Images were recorded on a 2k × 2k pixel charge-coupled device camera at a defocus level between –9 and –4 μm. The pixel size at the specimen level varied between 1 nm and 0.6 nm. The total electron dose per tilt-series was estimated at ~40 e Å^{–2}. Tomographic reconstructions were calculated by weighted back-projection and after a first rough manual segmentation they were refined with a combination of thresholding and erosion–dilation operations³³. Sub-tomograms were automatically extracted around the dense midline, aligned by their autocorrelation properties to common orientations, and classified with missing-wedge-compensated multivariate statistical analysis¹⁶. Semi-automated fitting was performed using the C-cadherin ectodomain structure¹⁴. To generate the artificial model, PDB 1L3W was superposed on a tetramer built from the interfaces in 1NCI and 1NCH¹⁵, and the angles of the monomers were manually changed to a rotation around the *x* axis of 1NCI of 30° and around the *z* axis of 8°. Detailed protocols, sample preparation, electron microscopy techniques, and image-processing methods are available in Supplementary Information.

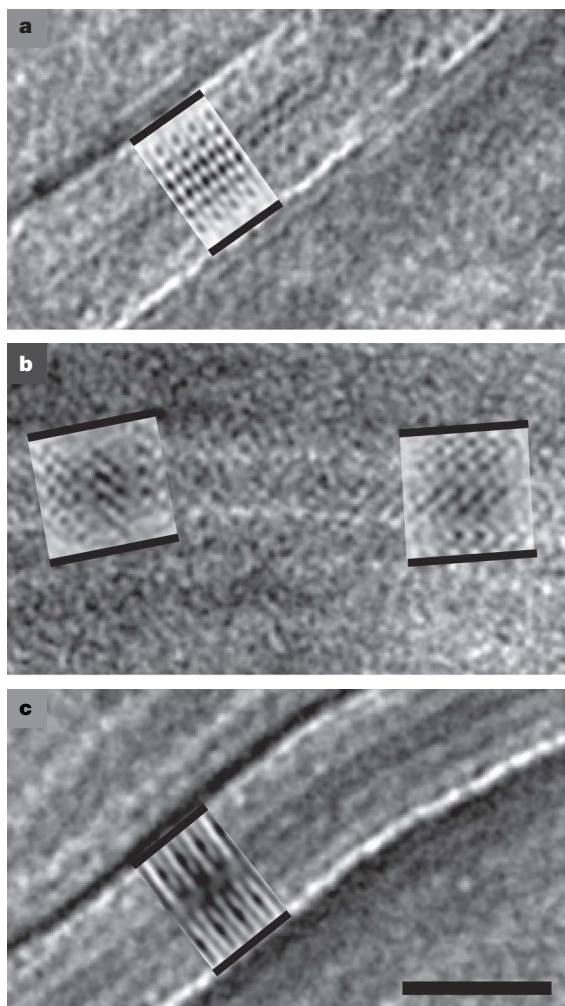


Figure 6 | Comparison of the modelled desmosome with 2D transmission CEMOVIS data of the skin. **a–c.** The projections of the model are indicated by black parallel lines and are superposed on the original CEMOVIS images. The desmosomes used for this comparison were chosen according to the same phenotype and were recorded at similar regions of the epidermis to our current data. The projection orientations between the three images correspond to an axial tilt of the model presented in Fig. 5 of 30° (**a**), 18° (**b**) and 58° (**c**). Scale bar, 35 nm.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 August; accepted 11 October 2007.

- Garrod, D. R., Merritt, A. J. & Nie, Z. Desmosomal adhesion: structural basis, molecular mechanism and regulation. *Mol. Membr. Biol.* **19**, 81–94 (2002).
- Patel, S. D., Chen, C. P., Bahna, F., Honig, B. & Shapiro, L. Cadherin-mediated cell–cell adhesion: sticking together as a family. *Curr. Opin. Struct. Biol.* **13**, 690–698 (2003).
- Nose, A., Tsuji, K. & Takeichi, M. Localization of specificity determining sites in cadherin cell adhesion molecules. *Cell* **61**, 147–155 (1990).
- Kottke, M. D., Delva, E. & Kowalczyk, A. P. The desmosome: cell science lessons from human diseases. *J. Cell Sci.* **119**, 797–806 (2006).
- Sali, A., Glaeser, R., Earnest, T. & Baumeister, W. From words to literature in structural proteomics. *Nature* **422**, 216–225 (2003).
- Al-Amoudi, A., Norlen, L. P. & Dubochet, J. Cryo-electron microscopy of vitreous sections of native biological cells and tissues. *J. Struct. Biol.* **148**, 131–135 (2004).
- Al-Amoudi, A., Dubochet, J. & Norlen, L. Nanostructure of the epidermal extracellular space as observed by cryo-electron microscopy of vitreous sections of human skin. *J. Invest. Dermatol.* **124**, 764–777 (2005).
- He, W., Cowin, P. & Stokes, D. L. Untangling desmosomal knots with electron tomography. *Science* **302**, 109–113 (2003).
- Hsieh, C. E., Leith, A., Mannella, C. A., Frank, J. & Marko, M. Towards high-resolution three-dimensional imaging of native mammalian tissue: electron tomography of frozen-hydrated rat liver sections. *J. Struct. Biol.* **153**, 1–13 (2006).
- Al-Amoudi, A. et al. Cryo-electron microscopy of vitreous sections. *EMBO J.* **23**, 3583–3588 (2004).
- Dubochet, J. et al. Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* **21**, 129–228 (1988).
- Forster, F., Medalia, O., Zauberman, N., Baumeister, W. & Fass, D. Retrovirus envelope protein complex structure *in situ* studied by cryo-electron tomography. *Proc. Natl Acad. Sci. USA* **102**, 4729–4734 (2005).
- Al-Amoudi, A., Studer, D. & Dubochet, J. Cutting artefacts and cutting process in vitreous sections for cryo-electron microscopy. *J. Struct. Biol.* **150**, 109–121 (2005).
- Boggon, T. J. et al. C-cadherin ectodomain structure and implications for cell adhesion mechanisms. *Science* **296**, 1308–1313 (2002).
- Shapiro, L. et al. Structural basis of cell–cell adhesion by cadherins. *Nature* **374**, 327–337 (1995).
- Forster, F., Pruggnaller, S., Seybert, A. & Frangakis, A. S. Classification of cryo-electron sub-tomograms using constrained correlation. *J. Struct. Biol.* (in the press).
- Shaikh, T. R., Hegerl, R. & Frank, J. An approach to examining model dependence in EM reconstructions using cross-validation. *J. Struct. Biol.* **142**, 301–310 (2003).
- Patel, S. D. et al. Type II cadherin ectodomain structures: implications for classical cadherin specificity. *Cell* **124**, 1255–1268 (2006).
- Brieher, W. M., Yap, A. S. & Gumbiner, B. M. Lateral dimerization is required for the homophilic binding activity of C-cadherin. *J. Cell Biol.* **135**, 487–496 (1996).
- Pertz, O. et al. A new crystal structure, Ca²⁺ dependence and mutational analysis reveal molecular details of E-cadherin homoassociation. *EMBO J.* **18**, 1738–1747 (1999).
- Takeda, H., Shimoyama, Y., Nagafuchi, A. & Hirohashi, S. E-cadherin functions as a cis-dimer at the cell–cell adhesive interface *in vivo*. *Nature Struct. Biol.* **6**, 310–312 (1999).
- Yap, A. S., Brieher, W. M., Pruschy, M. & Gumbiner, B. M. Lateral clustering of the adhesive ectodomain: a fundamental determinant of cadherin function. *Curr. Biol.* **7**, 308–315 (1997).
- Nollet, F., Kools, P. & van Roy, F. Phylogenetic analysis of the cadherin superfamily allows identification of six major subfamilies besides several solitary members. *J. Mol. Biol.* **299**, 551–572 (2000).
- Vestweber, D. & Kemler, R. Identification of a putative cell adhesion domain of uvomorulin. *EMBO J.* **4**, 3393–3398 (1985).
- Shan, W. S. et al. Functional cis-heterodimers of N- and R-cadherins. *J. Cell Biol.* **148**, 579–590 (2000).
- Tamura, K., Shan, W. S., Hendrickson, W. A., Colman, D. R. & Shapiro, L. Structure–function analysis of cell adhesion by neural (N-) cadherin. *Neuron* **20**, 1153–1163 (1998).
- Nagar, B., Overduin, M., Ikura, M. & Rini, J. M. Structural basis of calcium-induced E-cadherin rigidification and dimerization. *Nature* **380**, 360–364 (1996).
- Chappuis-Flament, S., Wong, E., Hicks, L. D., Kay, C. M. & Gumbiner, B. M. Multiple cadherin extracellular repeats mediate homophilic binding and adhesion. *J. Cell Biol.* **154**, 231–243 (2001).
- Sivasankar, S., Brieher, W., Lavrik, N., Gumbiner, B. & Leckband, D. Direct molecular force measurements of multiple adhesive interactions between cadherin ectodomains. *Proc. Natl Acad. Sci. USA* **96**, 11820–11824 (1999).
- Chen, C. P., Posy, S., Ben-Shaul, A., Shapiro, L. & Honig, B. H. Specificity of cell–cell adhesion by classical cadherins: critical role for low-affinity dimerization through β -strand swapping. *Proc. Natl Acad. Sci. USA* **102**, 8531–8536 (2005).
- Garrod, D. R., Berika, M. Y., Bardsley, W. F., Holmes, D. & Taberner, L. Hyperadhesion in desmosomes: its regulation in wound healing and possible relationship to cadherin crystal structure. *J. Cell Sci.* **118**, 5743–5754 (2005).
- Dubochet, J. & Sartori Blanc, N. The cell in absence of aggregation artifacts. *Micron* **32**, 91–99 (2001).
- Frangakis, A. S. & Hegerl, R. in *Electron Tomography* (ed. J. Frank) 353–370 (Springer, New York, 2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Saibil, B. Boettcher, A. Seybert and J. Dubochet for suggestions and for critically reading the manuscript. This work was supported by grants from the FP6 Marie Curie mobility network and EMBO fellowships to A.A.-A. and from the FP6 3DEM network of excellence to A.S.F.

Author Contributions A.A.-A. prepared the samples, and recorded and interpreted the data sets. D.C.D. developed algorithms for aligning and classifying the data. M.J.B. performed a structural bioinformatics analysis. A.S.F. analysed the data. A.A.-A. and A.S.F. wrote the manuscript.

Author Information The cadherin map has been deposited in the EBI Macromolecular Structure Database with accession number EMD-1374. The software is available at http://www-db.embl.de/jss/EmblGroupsHD/g_247?sP=7 or on request. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.S.F. (frangak@embl.de).

METHODS

Tissue preparation. The skin area for biopsy had not been exposed to any detergents, treatments or skincare products before experimentation. We immediately submerged the biopsies in 1-hexadecane (Merck) to avoid dehydration during preparation. We then used a punching tool to extract disks of the samples with a diameter of 1 mm and a thickness of approximately 100–150 μm . We placed the sample disks in the cavities (diameter 1.5 mm; depth 0.1 mm) of copper membrane carriers (Leica Microsystems), and filled the void space with 1-hexadecane. No cryoprotectants were used in these preparations.

Freezing. Within 90 s of their extraction, we cryo-fixed the skin biopsies in the EMPACT2 high-pressure freezer (Leica Microsystems), which reaches a pressure of 2000 bar within 15 ms and has a nominal cooling rate of approximately $20,000\text{ }^{\circ}\text{C s}^{-1}$. We then transferred the frozen samples to the cryo-chamber of a Leica EM FC6/UC6 ultramicrotome (Leica Microsystems) for cryo-sectioning.

Cryo-sectioning. We mounted and fixed the membrane carriers containing the frozen skin samples in the flat chuck of the Leica EM FC6 cryochamber. We trimmed the samples with a 20° trimming diamond knife (Diatome) and cryo-sectioned them at $-140\text{ }^{\circ}\text{C}$ with a nominal section thickness of 50 nm using a 35° diamond knife (Diatome) with a clearance angle of 6° . Cutting speed was set to 1.0 mm s^{-1} . We transferred the sections to quantifoil R3.5/1 300 mesh copper grids (Quantifoil Micro Tools) using an eyelash glued to a wooden stick, and then pressed them with a cryo-tool (Leica Microsystems) to ensure good attachment to the grid. While attached to the grid, the surfaces of some sections were labelled with quantum dots by dipping the grids for a few seconds in a solution containing 9.8 nm PbS quantum dots in Toluene (Evident Technologies) and diluted in isopentane (Merck) to a concentration of 4% (V/V), according to the protocol of ref. 34. We then rinsed the grids briefly with ethane to remove excess isopentane, which solidifies below $-160\text{ }^{\circ}\text{C}$, and stored them in liquid nitrogen. We used electron diffraction to examine the state of water in the cryo-sections. We chose good vitreous sections for data collection, and discarded non-vitrified specimens or morphologically bad samples.

Tilt-series collection. We collected single-axis tilt series of desmosomal regions of more than 50 cells covering an angular range from -65° to $+65^{\circ}$ with 1 or 2 degrees angular increment in a Tecnai G2 Polara microscope (FEI) cooled to liquid nitrogen temperature and equipped with a Gatan postcolumn GIF 2002 energy filter. We carried out data acquisition under low-dose conditions using the UCSF tomography software³⁵. Images were recorded on a $2\text{ k} \times 2\text{ k}$ pixel CCD camera at a defocus level between -9 and $-4\text{ }\mu\text{m}$, depending on the pixel size at the specimen level, which varied between 1 nm and 0.6 nm, respectively (Supplementary Table 2). The total electron dose was around $40\text{ e}^{-}\text{ }\text{\AA}^2$ per tilt-series.

Alignment and segmentation procedure. We aligned the projection images of the samples without quantum dots with respect to a common origin using cross-correlation techniques. The merit figure of the aligned series used for the reconstruction and subsequent processing was less than 0.6 nm. This indicates that the sample did not change during the recording process, and the tilt series were properly aligned. We performed the reconstructions using weighted back-projection algorithms and visualized them with isosurface and volume-rendering techniques in the Amira software package (Mercury Computer Systems, www.amiravis.com). We de-noised the three-dimensional (3D) images with nonlinear anisotropic diffusion³⁶ and semi-automatically segmented them using erosion and dilation operations³³ after a rough manual segmentation of the reconstructions. We set the isosurface thresholds manually so that the underlying structures were faithfully represented, or, when visualizing cadherin molecules, so that the thickness of the isosurface was $\sim 3\text{ nm}$.

Processing of the desmosomal regions. We were guided by the directive not to impose any template or model on the data. Because several models for the arrangement of the cadherin molecules exist in the literature, biasing the processing towards a particular model would almost certainly converge to the starting model, prohibiting proper interpretation of the results. Consequently, during the alignment of the sub-tomograms to a common origin, we avoided similarity measurement techniques (for example, cross-correlation) that are prone to model biasing. Instead, we developed the following procedure; we also used computational techniques to extract and process the information.

We extracted the desmosomal region, which can be identified owing to its phenotype, from the reconstruction. First, we computationally localized the dense midline between the two juxtaposed cell membranes. The dense midline appears dense only in the projection images; in the reconstruction it does not have a higher contrast than other protein regions. Thus, we low-pass-filtered the desmosomal region and then created a binary image after applying a threshold. In the binary image the cell membranes are clearly visible and their boundaries can be precisely estimated with labelling techniques (for a detailed description of the selection procedure, see ref. 33 and Supplementary Fig. 1). We approximated

the position of the dense midline as the shortest distance between the two opposing cell membranes. Through visual inspection the calculated positions correspond well to the dense midline.

At the calculated position of the dense midline, we chose pixels with low intensities (high density) and extracted regions around these positions to produce sub-tomograms. The minimum allowed distance between two selected pixels was varied between 4 and 10 nm without having a major impact on the results. Depending on the size of these sub-tomograms, one or several interacting dimers from the two juxtaposed cell surfaces could be included. The dimers are flexible and therefore large sub-tomograms did not align well. We found that sub-tomograms of 18 nm in diameter, which are capable of containing 9 putative dimers, gave the best trade-off between alignment quality and the area required to resolve the molecular architecture of the desmosomes.

We then aligned the sub-tomograms. We assumed that the translational alignment of the sub-tomograms was roughly correct, and first aligned them rotationally. For this, we exploited the properties of the auto-correlation functions of the sub-tomograms. Auto-correlation functions are used in various applications and tend to enhance inherent properties of the processed feature. To rotationally align the sub-tomograms, we calculated the auto-correlation function of each one (Supplementary Fig. 2). These show a pronounced elongation in three different directions: that of (1) the dense midline (Supplementary Fig. 2a); (2) the stapling of the cadherin molecules in the desmosome (Supplementary Fig. 2b); and (3) the interacting cadherin molecules spanning the space between opposing membranes (Supplementary Fig. 2a). All three directions can be precisely measured through the radon (or Hough) transformation in the corresponding plane and provide information about the following properties of each analysed sub-tomogram: the local orientation of the desmosome, which can be expressed in the three Euler angles; and the periodicity and orientation of the cadherin molecules with respect to the orientation of the desmosome (again in three Euler angles, especially in the case where the cadherin molecules are not perfectly perpendicular to the membrane). These two Euler triplets (the three Euler angles measured for the local orientation of the desmosome, and the three Euler angles for the local orientation of the cadherin molecules with respect to the plasma membrane) can be used to align all sub-tomograms with respect to each other, as well as to determine local flexibilities of the cadherin molecules within the desmosome. To achieve a full rotational alignment, the eigen-rotation of the molecule also needs to be determined. This, however, was not visible in the auto-correlation function and was calculated in a different way. After aligning the sub-tomograms around the two angles of the first Euler triplet, which could be determined through the auto-correlation function, we rotated the sub-tomograms around phi and projected them in a common direction. At specific angles, these showed a periodicity at $\sim 5\text{ nm}$ (similar to the CEMOVIS images), which was recognizable in both Fourier and real space. We chose the positions with the highest amplitudes in the Fourier space. With this procedure the sub-tomograms could be rotationally aligned according to their own properties, without imposing any template. The aligned sub-tomograms are presented in Supplementary Fig. 3. In the centre of the average of the aligned sub-tomograms is a dark spot; this is expected because all the sub-tomograms were cropped around pixels with low intensities, which are now overrepresented in the averaged structure. However, this dark spot disappears after the classification and translational alignment of the sub-tomograms (Fig. 3)¹⁶. The membranes of the juxtaposed cells and the cadherin molecules are also visible and show a periodicity. We refined the translational positions of the sub-tomograms by aligning the individual sub-tomograms with respect to their average within a range of a few pixels (± 2 pixels).

Classification procedure. The cadherin molecules in the desmosome show clear local periodicity and resemble a quasi two-dimensional (2D) crystal. Unfortunately, the crystal is not rigid enough for the 3D reconstruction to be reduced to a 2D crystallography problem. Between neighbouring molecules an angle variation of up to 20° can be measured. This variation limits the resolution and hinders the elimination of the missing wedge. Our aim was to visualize the architecture of the cadherin molecules, rather than to visualize the shape of the cadherin dimers, which has already been solved by X-ray crystallography¹⁴. We therefore had to include as many cadherin molecules as possible to resolve their organization while using as few as possible to achieve a resolution sufficient for unambiguous fitting. Improving the resolution on a single dimer and eliminating the missing wedge on the dimer or a tetramer results in the loss of contrast of the other neighbouring molecules and hence prohibits proper fitting. We found that a cylindrical mask with a diameter of 18 nm and a length of 34 nm, which includes several putative dimers but excludes the cell membranes, produced the best balance. We performed the classification of the sub-tomograms with multivariate statistical analysis and the KerDenSOM algorithm^{16,37}. In both algorithms the missing wedge was considered in order to improve the performance and to decouple the orientation of the particles from their classification assignment.

Both algorithms performed similarly and molecules were visible in the resulting classes (Fig. 3).

Fitting algorithm. We low-pass-filtered the cadherin molecules (PDB coordinates 1L3W¹⁴) to 3.4 nm and imposed on them the contrast transfer function of the microscope³⁸. The central cadherin molecule was fitted with a template-matching algorithm that was based on normalized cross-correlation. The positions of the remaining molecules were then constrained by the positions of the previously fitted molecules, which were essentially used as spacers. This procedure allowed us to fit the densities of the molecules reliably, even far from the centre of the sub-tomogram, and to partially overlap density gaps resulting from the missing wedge and the contrast transfer function of the microscope. Finally, we imposed an additional constraint on the fitting procedure: Cadherin molecules from opposing cell surfaces were forced to attach to their partner molecules from the tryptophan side, as seen in the X-ray structure of C-cadherin by ref. 14. Because the present resolution of the sub-tomograms can only visualize the general arrangement of the cadherin molecules and does not allow for the interpretation of the interaction site, this is a justified approximation. The resulting fitted molecules resemble very well the densities of the sub-tomograms, and visualize clearly the arrangement of the cadherin molecules.

Generation of the model. To generate an artificial model based on our findings, we superposed four copies of PDB 1L3W¹⁴ on to a tetramer built from 1NCI and 1NCH^{14,15}. The angles between the monomers did not resemble our data sufficiently well, so we manually changed them by a rotation around the *x*-axis of 1NCI of 30° and around the *z*-axis of 8°. This arrangement resembled the fitted data as a tetramer (two *trans* and one *cis* interaction) with the best score (Fig. 5). From the periodic arrangement of the cadherin molecules, we estimated their spacing using the power spectrum. This closely resembled previous findings from the X-ray structures, and was used to complete the full cadherin array.

34. Masich, S., Ostberg, T., Norlen, L., Shupliakov, O. & Daneholt, B. A procedure to deposit fiducial markers on vitreous cryo-sections for cellular tomography. *J. Struct. Biol.* **156**, 461–468 (2006).
35. Zheng, Q. S., Braunfeld, M. B., Sedat, J. W. & Agard, D. A. An improved strategy for automated electron microscopic tomography. *J. Struct. Biol.* **147**, 91–101 (2004).
36. Frangakis, A. S. & Hegerl, R. Noise reduction in electron tomographic reconstructions using nonlinear anisotropic diffusion. *J. Struct. Biol.* **135**, 239–250 (2001).
37. Pascual-Montano, A. *et al.* A novel neural network technique for analysis and classification of EM single-particle images. *J. Struct. Biol.* **133**, 233–245 (2001).
38. Frangakis, A. S. *et al.* Identification of macromolecular complexes in cryoelectron tomograms of phantom cells. *Proc. Natl Acad. Sci. USA* **99**, 14153–14158 (2002).

ARTICLES

Intrinsic motions along an enzymatic reaction trajectory

Katherine A. Henzler-Wildman¹, Vu Thai¹, Ming Lei¹, Maria Ott³, Magnus Wolf-Watz^{1†}, Tim Fenn^{2‡}, Ed Pozharski^{2‡}, Mark A. Wilson^{2‡}, Gregory A. Petsko², Martin Karplus^{4,5}, Christian G. Hübner^{3†} & Dorothee Kern¹

The mechanisms by which enzymes achieve extraordinary rate acceleration and specificity have long been of key interest in biochemistry. It is generally recognized that substrate binding coupled to conformational changes of the substrate–enzyme complex aligns the reactive groups in an optimal environment for efficient chemistry. Although chemical mechanisms have been elucidated for many enzymes, the question of how enzymes achieve the catalytically competent state has only recently become approachable by experiment and computation. Here we show crystallographic evidence for conformational substates along the trajectory towards the catalytically competent ‘closed’ state in the ligand-free form of the enzyme adenylate kinase. Molecular dynamics simulations indicate that these partially closed conformations are sampled in nanoseconds, whereas nuclear magnetic resonance and single-molecule fluorescence resonance energy transfer reveal rare sampling of a fully closed conformation occurring on the microsecond-to-millisecond timescale. Thus, the larger-scale motions in substrate-free adenylate kinase are not random, but preferentially follow the pathways that create the configuration capable of proficient chemistry. Such preferred directionality, encoded in the fold, may contribute to catalysis in many enzymes.

A folded protein is not a unique structure, but includes an ensemble of folded substates that are sampled at physiological temperatures^{1–4}. As pointed out about 40 years ago⁵, protein folding cannot proceed by random sampling of all possible conformations. In analogy to the energy landscape view of the folding process^{6,7}, rearrangements within a folded protein are similarly directed by the energy landscape^{1,4}. Although the lowest energy structures can often be determined experimentally, an understanding of other conformations and the transitions among them is still in its infancy.

Synergy between structure and plasticity results in the unique power of biocatalysts (enzymes). The chemical mechanisms of many enzymatic reactions are known in great detail thanks to advances in classical enzymology and structural biology. For a number of enzymes, snapshots of conformations that are sampled during catalysis have been obtained using ligands, substrates and inhibitors. Recently, transitions between these states have been measured by nuclear magnetic resonance (NMR) relaxation experiments with substrate analogues^{8–11} or during catalysis^{12,13}, as well as by single-molecule fluorescence resonance energy transfer (FRET)^{14–19}. Here we explore how an enzyme, adenylate kinase, reaches a catalytically competent conformation in which the reactive groups are brought into close proximity in a position favouring catalysis. Using X-ray crystallography, NMR, single-molecule FRET, normal mode analysis (NMA) and molecular dynamics simulations, we identify conformational substates along this reaction trajectory that are already present in the substrate-free enzyme, and determine timescales for the transitions among them.

Open and closed conformation of *Aquifex* adenylate kinase

The crystal structure of adenylate kinase (Adk) from the hyperthermophile *Aquifex aeolicus* (*Aquifex* Adk) has been solved (Fig. 1). Adks are enzymes that maintain the cellular equilibrium concentration of

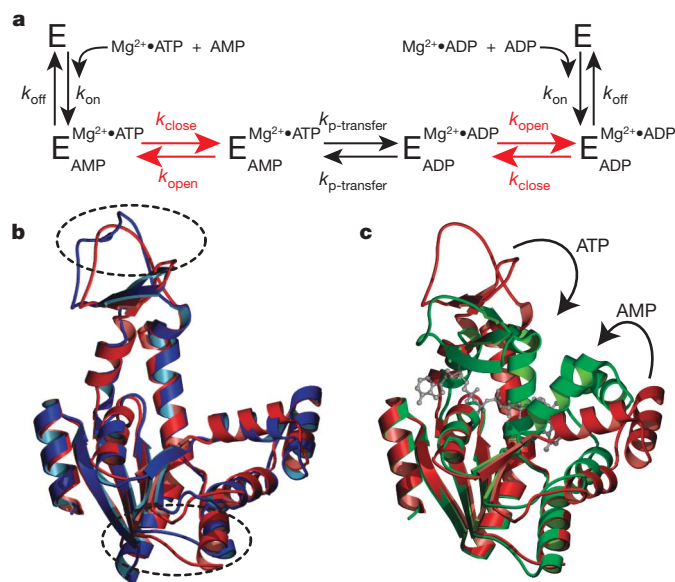


Figure 1 | Kinetic model and X-ray structure of *Aquifex* Adk. **a**, Proposed reaction scheme for the enzyme adenylate kinase (E) including the steps of substrate binding (k_{on}), lid closing (k_{close}), phospho-transfer ($k_{\text{p-transfer}}$), lid opening (k_{open}) and substrate dissociation (k_{off}). **b**, Superposition of molecule A (for definition, see Fig. 2) of apo *Aquifex* (red) with apo *E. coli* (blue) Adk reveals only small changes in the overall structure between the homologues, as indicated by dashed ovals. **c**, Superposition of apo *Aquifex* Adk (red) and *Aquifex* Adk in complex (green) with the substrate analogue Zn^{2+} -Ap5A (shown as ball and stick in grey) demonstrates the closure of the ATP and AMP lids on substrate binding.

¹Department of Biochemistry and Howard Hughes Medical Institute, ²Department of Biochemistry, Brandeis University, Waltham, Massachusetts 02454, USA. ³Institute of Physics, Martin Luther-University Halle-Wittenberg, D-06120 Halle, Germany. ⁴Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ⁵Laboratoire de Chimie Biophysique, ISIS, Université Louis Pasteur, F-67000 Strasbourg, France. †Present addresses: University of Umeå, Department of Chemistry, SE-90187 Umeå, Sweden (M.W.-W.); Departments of Molecular and Cellular Physiology and Howard Hughes Medical Institute, Stanford University, Stanford, California 94305, USA (T.F.); Department of Pharmaceutical Sciences, University of Maryland, Baltimore, Maryland 21201, USA (E.P.); Department of Biochemistry and the Redox Biology Center, University of Nebraska, Lincoln, Nebraska 68588, USA (M.A.W.); University at Lübeck, Institute of Physics, 23538 Lübeck, Germany (C.G.H.).

adenylate nucleotides by catalysing the reversible conversion of ATP and AMP into two ADP molecules (Fig. 1a). Adks are homologous in sequence and structure²⁰, and typically have a core domain plus ATP and AMP lids. The overall structure of *Aquifex* Adk is strikingly similar to that of *Escherichia coli* Adk²¹. The most obvious differences are shortening of loops in the ATP lid and at the base of the AMP lid, and the presence of five additional prolines in *Aquifex* Adk²².

To explore conformational rearrangements sampled during catalysis, we solved the structure of *Aquifex* Adk complexed with the bi-substrate analogue Zn^{2+} •Ap5A (Fig. 1c). Ap5A consists of two adenosine nucleosides covalently linked through a pentaphosphate. It binds tightly to Adk, mimicking the ternary complex²³. For phosphotransfer, both lids must close to exclude bulk water from the active site and to bring the substrates into position for phosphotransfer. Consequently, large conformational changes are observed on binding of Ap5A to *Aquifex* Adk that are very similar to the changes observed in other Adks in complex with Mg^{2+} •Ap5A²³ or other substrates or analogues²⁰. A large number of X-ray structures of nucleoside monophosphate kinases with and without ligands were used previously to create a movie of ligand-induced conformational changes²⁰.

Substates in apo Adk along the reaction trajectory

Although those results would fit nicely with the standard view of ligand-induced conformational change between single apo and ligand-bound states, the combined crystallographic, NMR, single-molecule and computational studies demand a fundamentally different picture. The asymmetric unit of apo *Aquifex* Adk consists of three molecules (A, B and C) with significantly different conformations (Fig. 2). Stabilization of these three particular substates seems to be caused by crystal contacts (see Supplementary Figs 1 and 2). Although this phenomenon is known for other crystal structures (for examples, see refs 24–27), different conformations are more frequently observed in crystals of molecules with and without bound ligands.

Are these substates relevant for catalysis? A least-squares superposition of the C α atoms of these three molecules with the Zn^{2+} •Ap5A-bound ‘fully closed’ form shows that these lid substates lie along a trajectory between the fully open and fully closed conformations (Fig. 2a). The substates of ligand-free Adk arise from hinge-bending motions of the lids relative to the core; the conformations within each domain are very similar among the substates (Fig. 2). Further bending of the same hinges results in full closure on inhibitor binding. These hinges have unique physical properties²² and are located at very similar positions in *E. coli* Adk^{21,28,29}. Although the hinges are found mostly in loop regions, one hinge for the ATP lid is located in the middle of a long α helix (hinge 8), which actually breaks into two helices on closure. The crystallographic results show the directionality of the rearrangements, but do not determine the probability of attaining these structures or the rates of transitions among them in solution.

Motions along the reaction trajectory measured by NMR

Previously reported snapshots of a series of conformations of Adks with and without bound nucleotides have been shown to lie on a calculated path between the open and closed state^{20,29}. Moreover, the transition from the closed to the open state has been shown recently to limit the overall catalytic rate¹³.

To address the question of whether the conformational sampling seen in the crystal structure also occurs in solution, ¹⁵N NMR relaxation dispersion experiments were performed on apo *Aquifex* Adk (Fig. 3a, c). These experiments allow detection of residues that undergo changes in their local chemical environments and characterization of the kinetics¹⁰. In an NMR experiment, conformational exchange in the microsecond-to-millisecond timescale causes increased dephasing of coherence, resulting in additional line-broadening of NMR signals by an amount (R_{ex}) that contributes to the measured overall transverse

relaxation rate (R_2^{eff}). R_{ex} can be refocused by increasing the power of an applied radio-frequency field (ν_{CPMG} , where CPMG stands for Carr–Purcell–Meiboom–Gill). From the dependence of R_{ex} on ν_{CPMG} explored in the relaxation dispersion experiment (Fig. 3a), three physical parameters for a two-site dynamic process can, in principle, be obtained: rates of interconversion (k_{ex}), relative populations of the exchanging species (p_{A} and p_{B}) and chemical shifts between the exchanging species ($\Delta\omega$) (see Methods).

Residues located in the hinge regions and in parts of both lids that show conformational heterogeneity in the crystal also show chemical exchange in solution (Fig. 3c). In these areas, fluctuations among the crystallographic substates would cause changes in the local environment, and consequently detectable chemical exchange. In contrast, nuclei in regions where motion does not cause any changes in their electronic environment, such as within a rigid domain, will not display R_{ex} in this experiment. We stress that the measured R_{ex} , manifested in the amplitude of the dispersion profiles, is related directly to $\Delta\omega$. Spatial information cannot be inferred from these data, but the kinetics of the fluctuations can be measured.

To quantitatively compare the kinetics of lid movement in apo *Aquifex* Adk with the kinetics during turnover¹³, ¹⁵N relaxation experiments were performed under the same physical conditions at 20 °C. However, the rates were too fast in apo *Aquifex* Adk (rates of 2,000–7,000 s^{−1} are compatible with the dispersion data, see Supplementary Fig. 6) for accurate analysis. Therefore, the experiments were repeated at 10 °C with perdeuterated protein. The lower temperature brings the rates into a time regime that is well determined

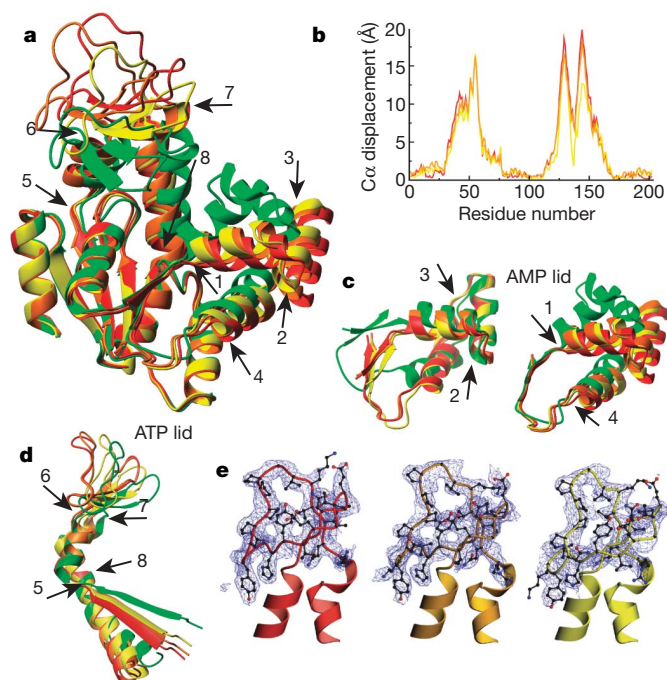


Figure 2 | Conformational substates of ligand-free *Aquifex* Adk detected in the crystal structure. **a**, Superposition of the three molecules, A, B and C, in the asymmetric unit of apo Adk (red, orange and yellow, respectively) and Adk complexed with Zn^{2+} •Ap5A (green; Zn^{2+} from the crystallization mother liquor is bound to the Mg^{2+} site). The substates A, B and C lie along the reaction trajectory towards the closed state. **b**, Backbone displacement of A (red), B (orange) and C (yellow) relative to the inhibitor-bound form. **c**, **d**, The conformational substates A, B and C are a result of motions around eight hinges, indicated by arrows (for details about the hinges, see ref. 22). For better visualization of the hinges of the ATP lid, the latter was rotated by 90° (**d**) with respect to **a**. The two views in **c** show the AMP lid with different segments overlaid to highlight the two distinct hinge pairs. **e**, $2F_o - F_c$ maps contoured at 1.0σ of the ATP lids of molecules A (left), B (centre) and C (right) show the quality of the electron density.

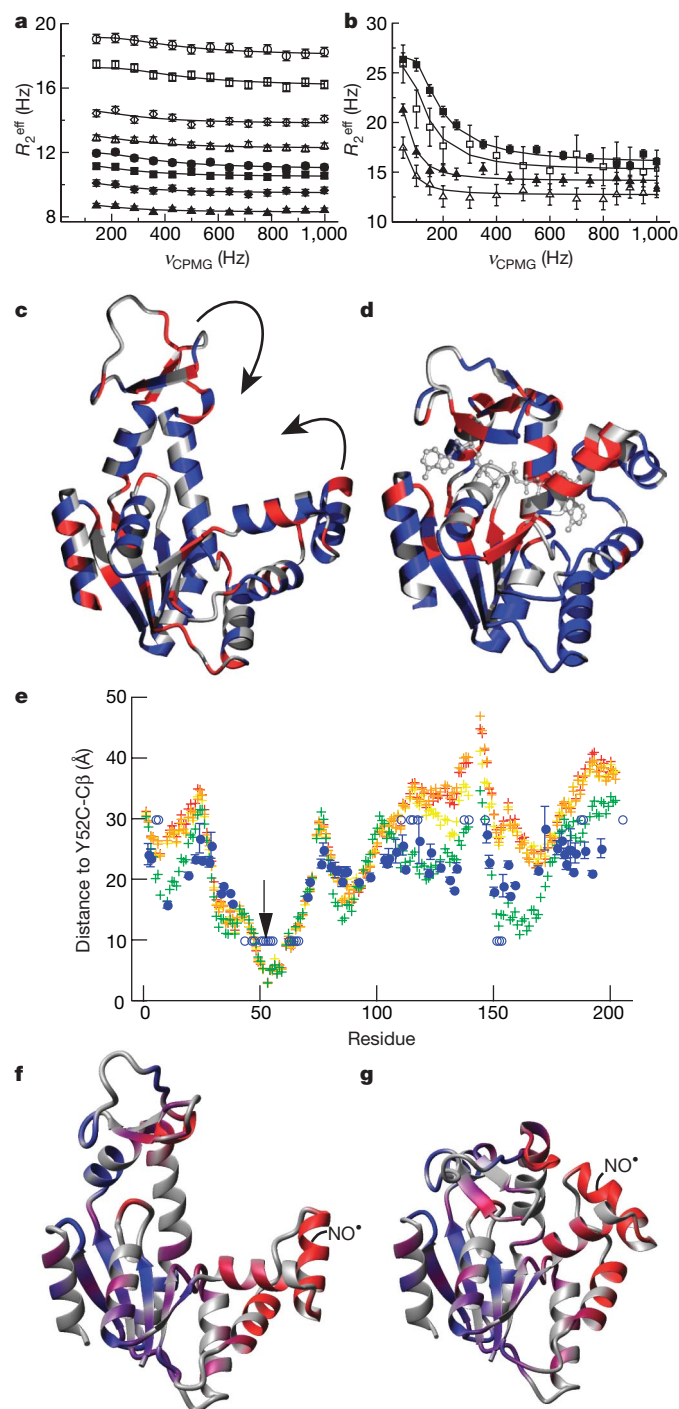


Figure 3 | Characterization of millisecond dynamics of *Aquifex* Adk in solution by NMR. **a–d**, The dependence of the ^{15}N effective transverse relaxation time (R_2^{eff}) on an applied radio-frequency field (ν_{CPMG})⁴⁶ for representative residues of $^2\text{H}/^{15}\text{N}$ apo Adk at 10 °C (**a**) and $^1\text{H}/^{15}\text{N}$ Adk in complex with Mg^{2+} •Ap5A at 20 °C (**b**) with errors bars (± 1 s.d.) based on duplicates and signal-to-noise ratios. Residues with conformational exchange (red), without exchange (blue) and overlapped or proline residues (grey) are mapped onto the crystal structure of apo Adk (**c**) and Adk complexed with Mg^{2+} •Ap5A (**d**). **e–g**, PRE by a spin label (NO^+) at residue 52 indicates that ligand-free Adk samples conformations resembling the closed state. **e**, Distances calculated from Y52C-C β to each amide proton for molecule A (red), B (orange) and C (gold) of apo Adk and inhibitor-bound Adk (green) are compared to distances calculated from the spin-label-induced line-broadening for ligand-free Adk (blue, error bars are ± 1 s.d.). Open circles represent H_N that are broadened beyond detection owing to the close proximity to the spin label (shown at 10 Å) or have no observable line broadening (shown at 30 Å). The PRE-derived distances are plotted onto the structure of the open (**f**) and closed (**g**) conformation as a continuous colour scale from blue (far away) to red (close proximity).

by CPMG relaxation experiments, whereas deuteration decreases the rate of transverse relaxation, resulting in a larger contribution of conformational exchange relative to the overall relaxation rate. Under these conditions, most of the exchanging residues can be fit with a single exchange rate of $950 \pm 150 \text{ s}^{-1}$ (mean \pm standard deviation (s.d.)) (Fig. 3a). Although this rate is well determined, the small amplitude of the dispersion profiles hampered characterization of the nature of the conformational transition.

The combination of crystallographic and NMR data indicate that the apo enzyme consists of an ensemble of conformations along the lid-motion trajectory required for catalytic function. Is a similar physical picture applicable for the inhibitor-bound, closed state? The X-ray structure shows one conformation (Fig. 1c). In contrast, ^{15}N relaxation dispersion experiments on Mg^{2+} •Ap5A-bound *Aquifex* Adk at 20 °C reveal conformational fluctuations with a rate of $300 \pm 150 \text{ s}^{-1}$ (mean \pm s.d.) that can be rationalized by a full opening/closing transition (Fig. 3b, d). Larger $\Delta\omega$ values in the presence of the ligand result in larger R_{ex} values, allowing a detailed analysis similar to the analysis during catalysis¹³. First, $\Delta\omega$ values calculated from the dynamics experiments are in agreement with the chemical shift differences between open and closed conformations (Supplementary Table 4 and Supplementary Fig. 4). Second, the calculated relative populations are highly skewed towards the closed state ($95 \pm 2\%$). These results imply that the enzyme fluctuates between the closed and open conformation, even in the presence of the inhibitor Mg^{2+} •Ap5A, with the equilibrium shifted far towards the closed form. It is not surprising that the rates measured for Adk bound to Mg^{2+} •Ap5A differ from those measured in the presence of natural substrates ($\sim 1,600 \text{ s}^{-1}$)¹³ because in Ap5A both nucleotides are covalently connected, significantly altering the energy landscape.

Motions along the reaction trajectory by molecular dynamics

The experiments described above show that ligand-free Adk can exist in multiple conformations with variable degrees of closure, and that transitions occur in solution on a millisecond timescale. Computational methods provide a means to link these spatial and dynamic characteristics. First, we performed molecular dynamics simulations in explicit water (individual water molecules included in the simulation) starting from the three crystallographic substates of apo *Aquifex* Adk: A, B and C. The largest root-mean-square fluctuations are located in the lid regions, with similar motional amplitudes for all three substates (Fig. 4a). Essentially the same conformational space is sampled in 10 ns starting from any of the three conformations (Fig. 4c and Supplementary Fig. 9). Because a 10 ns simulation is long enough for one molecule to visit all three crystallographic conformations, the energy barriers for interconversion among these substates must be low. Furthermore, these simulations allow investigation of the correlations between the motions of the two lids, which cannot be measured in an ensemble-averaged experiment. From the time traces (Fig. 4b) and a covariance matrix calculated after superimposing the core domain of all snapshots (Supplementary Fig. 8), it is apparent that the two lids move independently, at least for the subspace sampled by the apo structure in 10 ns.

To address the question of whether the lid fluctuations are random or have preferred directionality, we performed a principal component analysis (PCA)³⁰ and a NMA³¹. Both methods calculate the directionality and frequency of motions. The lowest frequency modes are often correlated with experimentally observed conformational changes^{29,32,33}. For *Aquifex* Adk, comparison of the vector connecting the most open (apo molecule A) and closed (inhibitor-bound) structures with the calculated lowest frequency modes shows that the first and second modes of the PCA contribute 49% and 63%, respectively, to the functionally important transition between the open and the closed states; this is consistent with results from a PCA of *E. coli* Adk³⁴. NMA provides similar results, with 80% contribution of its

first mode to the experimentally observed direction of lid closure (Fig. 4d, e and Supplementary Fig. 10).

Notably, the three 'open substates', A, B and C, are virtually along the same conformational coordinate, consistent with hinge-bending accounting for the observed conformational transitions. Most of these hinges have been suggested to be responsible for lid closure on the basis of structural comparison of the open and closed conformations of a variety of mesophilic adenylate kinases^{21,28}. Recent computational pathway analysis of ligand-free *E. coli* Adk has also identified changes in these regions along the pathway^{29,33}. Importantly, our crystal structures provide experimental evidence for bending at all hinges in ligand-free *Aquifex* Adk. Moreover, we show that lid displacements observed in the crystal of apo Adk are not separated by large free energy barriers, whereas there is a significantly larger barrier to reach the fully closed state (NMR and single-molecule FRET experiments). These results are in agreement with the main barrier being closer to the closed conformation, as indicated by the pathway analysis^{29,33}.

Sampling of conformational substates by single-molecule FRET

Whereas the conformational substates seen in the X-ray structure of apo Adk were sampled in 10-ns molecular dynamics simulations, the NMR relaxation experiments were indicative of lid motions in the

millisecond time regime. However, the NMR data did not provide information about the amplitude of motion. To address this difference in timescale between molecular dynamics simulations and NMR, we performed single-molecule FRET experiments³⁵. *Aquifex* Adk was fluorescently labelled using Alexa488 as the donor and Alexa633 as the acceptor in the AMP and ATP lids. The dyes (calculated Förster radius, R_0 , of 48 Å) and labelling sites were selected to optimally detect lid–lid distance changes going from the most open to the fully closed state (Fig. 5a). The enzymatic activity of the labelled enzyme is 35% of that of the wild type³⁶.

First we measured single-molecule FRET efficiency (E_t) distributions of both the ligand-free and the Mg^{2+} •Ap5A-bound form diffusing freely in solution (Fig. 5b, c)³⁷. FRET efficiencies were calculated from the total photon counts emitted by the donor and acceptor during the dwell time in the focus of the laser. The Mg^{2+} •Ap5A-bound form was used as a reference point for calibration of relative FRET efficiencies with respect to distance¹⁸. On the basis of our NMR data, the Mg^{2+} •Ap5A-bound form is highly skewed towards the fully closed conformation (Fig. 3b, d). This result is clearly reinforced by single-molecule FRET, which has the major population centred at $E_t = 0.83$ (corresponding to the closed state) with a tail at lower E_t values (open states) (Fig. 5c).

Interestingly, a similar closed state is sampled to a significant fraction even in the absence of any ligand (Fig. 5b). However, most of the ligand-free molecules have lower FRET efficiencies corresponding to conformations with increased lid–lid distances. The width of the distributions is caused by shot noise (see Methods), and the peak values correspond to lid–lid distances of the fully open and closed conformations. Two key conclusions can be drawn. First, apo Adk seems to sample both open and closed states in solution. Second, the timescale of interconversion between the open and closed conformations must be on the order of the residence time of the molecules in the focus (about 0.1 ms, as determined by fluorescence correlation spectroscopy, data not shown) or slower, because complete averaging of the FRET efficiencies would occur for much faster internal motions, resulting in a single E_t distribution³⁸.

To quantify the rates of lid opening and closing, transient fluorescence intensities of both the donor and the acceptor were recorded for individual Adk molecules immobilized on the surface of the microscope coverslip (Fig. 5d). The immobilized Adk molecules have full catalytic activity as probed by a luciferase assay (Supplementary Fig. 13). Indeed, opening and closing events could be detected directly for ligand-free Adk using 200- μs bin times. The similarity of the shape of the E_t histogram from the time traces (Fig. 5d) and the diffusion experiment (Fig. 5b) is a nice demonstration of ergodicity of the system. Instead of setting a unique threshold value for discriminating closed and open states, an 'ambiguous zone' was defined for E_t values where the distributions of the open and closed states overlap (Fig. 5b–d). A change of state was assigned only when the ambiguous zone was crossed by E_t values. The lifetime distributions of open and closed states followed exponential statistics (Fig. 5e, f). Because the fitted rates were similar to the experimental time resolution it was apparent that those raw rates had to be corrected for missed events (see Methods), yielding opening and closing rates of $6,500 \pm 500 \text{ s}^{-1}$ and $2,000 \pm 200 \text{ s}^{-1}$ (mean \pm s.d.), respectively.

Clear evidence for our interpretation of opening/closing comes from a correlation analysis exhibiting pronounced anti-correlation in the cross-correlation function (Fig. 5g, details in Methods). The overall rate of these fluctuations is $7,000 \pm 2,000 \text{ s}^{-1}$ (mean \pm s.d.), in very good agreement with the sum of opening and closing rates obtained from the E_t -state analysis above and from NMR experiments at 20 °C ($k_{\text{open}} + k_{\text{close}}$ between 2,000 and 7,000 s^{-1} , Supplementary Fig. 6). Similar time-resolved single-molecule FRET experiments were performed on Mg^{2+} •Ap5A-bound Adk, showing slower opening/closing transitions; this is in agreement with the kinetics measured by NMR (Fig. 3b and Supplementary Figs 16 and 17). It is important to note that we are using an oversimplified two-state model. In reality,

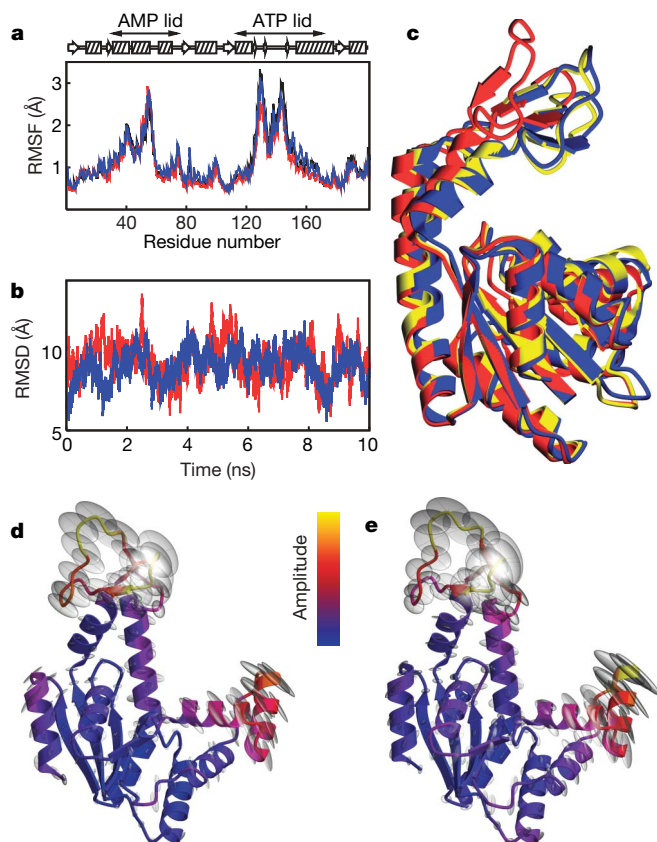


Figure 4 | Dynamics of *Aquifex* Adk computed by molecular dynamics simulations and normal mode analysis. **a**, The root mean square fluctuations (RMSF) are shown from 10-ns molecular dynamics simulations of molecule A (blue), B (red) and C (black) in explicit water. **b**, The corresponding time traces of the ATP lid (red) and AMP lid (blue) are visualized as root mean square deviation (RMSD) between simulated snapshots (starting from molecule A) and the crystal structure of the closed state. **c**, One snapshot (blue) from the molecular dynamics trajectory of molecule A is superimposed onto the crystal structures of molecule A (red) and molecule C (yellow). Relative amplitudes and directions of the ten combined lowest frequency modes calculated by PCA³⁰ (**d**) and NMA³¹ (**e**) are visualized by the direction and size of the silver ellipsoids together with the colour scale.

more states are sampled, such as single-lid-closed conformations²⁰. It is possible that these single-lid-closed conformations are responsible for the smaller ratio between the open and closed populations determined by the single-molecule FRET experiments relative to the more skewed populations estimated from NMR experiments. These questions will be addressed in future experiments using three-colour FRET on triple-labelled Adk.

Sampling of conformational substates by NMR PRE

To corroborate our finding that ligand-free Adk in solution samples a substate similar to the fully closed inhibitor-bound crystal structure, we measured paramagnetic NMR relaxation enhancement effects (PRE)³⁹ from a spin label attached to the AMP lid (Fig. 3e–g). An unpaired electron increases the relaxation rate of other nuclei in an r^{-6} distance-dependent manner³⁹. Notably, severe line broadening

was observed for residues in the ATP lid that are far away from the spin-label attachment point in the open substates, in addition to the expected large effect on residues that are close to the spin-label site. However, all residues with line-broadening effects are clustered within a 30 Å radius of the spin label when mapped on the fully closed conformation (Fig. 3g). In an ensemble of conformations, short distances will dominate the observed PRE effects owing to the r^{-6} -distance-dependence, thus biasing the distance distribution. Therefore, the close agreement of the calculated PRE distances with the ones from the closed state should not be misinterpreted as a high population of the closed state for ligand-free Adk. The PRE data do not provide rates for the conformational transitions or populations. However, they unambiguously demonstrate sampling of states structurally similar to the closed state because distances from the spin label to all amides within 30 Å are detected. This method is complementary to the FRET method, which directly measures rates and populations but only a single distance.

Discussion

For catalysis, defined conformational changes pre-organize active-site residues in the configuration needed to facilitate chemistry^{40–42}. Random sampling of all possible folded configurations would degrade catalytic efficiency, a situation analogous to protein folding, which is similarly biased so as to sample only a small portion of the energy landscape. This concept of a hierarchy of substates has been pioneered by extensive studies on myoglobin^{1,3,4}. Here we show that motions in apo Adk occur preferentially in the direction of the catalytically competent closed conformation. The NMR relaxation dispersion, single-molecule FRET and paramagnetic relaxation enhancement experiments reveal that in the absence of substrate, Adk samples a minor state that is comparable to the closed state observed during catalysis. The concept of preferred fluctuations to states resembling the substrate-bound states directed by the energy landscape has been discussed recently⁴³. Although the large amplitude conformational transition in Adk is a relatively slow (microsecond-to-millisecond), infrequent event, faster (nanosecond) and more frequent transitions of smaller amplitude along the same trajectory are revealed by combining X-ray crystallography, molecular dynamics simulations and normal mode calculations. This hierarchy in space and time indicates that slow conformational transitions occurring on the timescale of catalytic turnover are facilitated by individual high-frequency local fluctuations^{2,22,44}. Biological function is the property selected by evolution. Thus, optimization of the energy landscape by selection of the substates and pathways needed for function is an element of efficient catalysis at physiological temperatures.

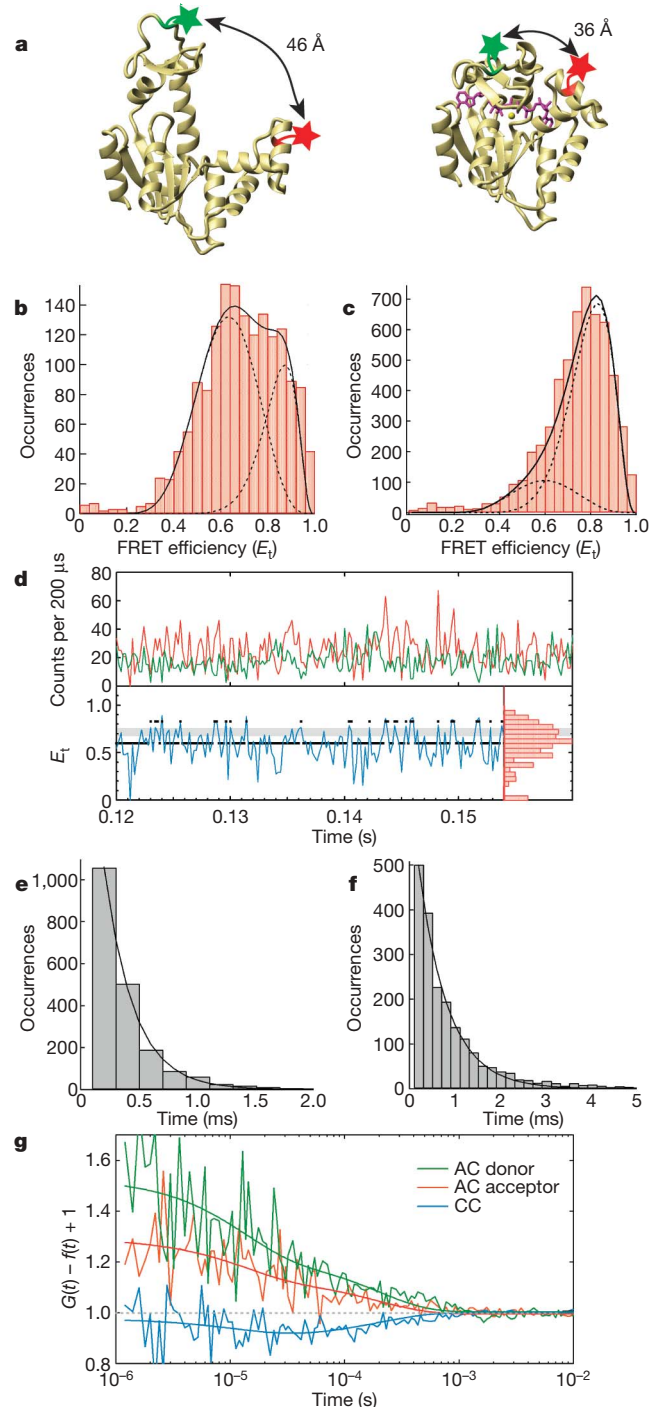


Figure 5 | Opening and closing of *Aquifex* Adk in solution by single-molecule FRET. **a**, Positions of fluorescent dyes Alexa488 (green) and Alexa633 (red) on Adk at residue 52 (AMP lid, right) and 145 (ATP lid, top) or vice versa. **b**, **c**, Histograms of FRET efficiencies of ligand-free (**b**) and Mg^{2+} -bound (**c**) Adk measured from single-molecule diffusion experiments. The dashed lines represent gaussian fits to the corresponding distance histograms back-transformed into FRET efficiency (E_t), and the solid line is the sum of those distributions. **d**, Section of a single-molecule fluorescence time trace of ligand-free Adk tethered on a glass surface. Donor (green) and acceptor (red) intensities are shown together with the corresponding E_t (blue), including the E_t histogram over 25 time traces (red, right). Lifetimes of the open and closed state (black lines) were determined on the basis of a transition zone for E_t values, where the gaussian distributions of the open and closed states overlap (grey). **e**, **f**, The resulting lifetimes of 25 time traces were averaged, fitted exponentially and corrected for missed events and triplet-state dynamics to yield $k_{open} = 6,500 \pm 500 \text{ s}^{-1}$ (**e**, mean \pm s.d.) and $k_{close} = 2,000 \pm 200 \text{ s}^{-1}$ (**f**), respectively. **g**, The intensity correlation function, $G(t)$, for auto-correlation (AC) and cross-correlation (CC) analysis averaged over six immobilized ligand-free Adk molecules. Anti-correlation of the cross-correlation function is observed in the 10^{-4} to 10^{-3} s time region with a fitted overall opening/closing rate of $7,000 \pm 2,000 \text{ s}^{-1}$ (mean \pm s.d.).

METHODS SUMMARY

Aquifex Adk was expressed and purified as described¹³. Fluorescently labelled Adk was prepared using Y52C and V145C mutants for labelling with Alexa633-maleimide and Alexa488-maleimide, and a 6×His tag on the carboxy terminus for tethering to a surface. Spin-labelled Adk was prepared using ¹⁵N-Adk(Y52C), which was reacted with (1-oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) methanethiosulfonate (MTSL) for 12 hours at 25 °C. Diffraction data of crystals of apo Adk and of Adk complexed with Zn²⁺•Ap5A were collected both at room temperature and at cryogenic temperature. Translation-libration-screw (TLS)⁴⁵ refinement was performed, dividing the three molecules into the core, the ATP lid and the AMP lid. Constant time ¹⁵N backbone CPMG relaxation dispersion experiments⁴⁶ were acquired in an interleaved manner incorporating transverse relaxation optimized spectroscopy (TROSY) selection⁴⁷ and fit as described¹². The effect of the spin label was determined by comparing the peak intensity in a pair of heteronuclear single quantum correlation (HSQC) spectra taken on the paramagnetic and diamagnetic form. The data were interpreted using ¹H R₂ values measured on the diamagnetic sample and converted to distances assuming a correlation time equal to the global tumbling time³⁹, measured for the unlabelled protein (τ_c)²². Molecular dynamics simulations were performed using the program CHARMM⁴⁸. Single-molecule FRET measurements were performed using a home-built sample-scanning optical confocal microscope. The transfer efficiency E_t for each burst was calculated according to the equation:

$$E_t = (I_A)/(I_A + I_D)$$

where I_D and I_A are the corrected intensity of the donor and acceptor channel, respectively, on donor excitation. Transient intensities on immobilized proteins were recorded with an acquisition time of 100 ns. The intensity time trace of a tethered molecule was taken by integrating photon counts for donor and acceptor channels with bin times between 200 μs and 1 ms. Auto- and cross-correlation functions were calculated from unbinned data for each individual molecule, subsequently averaged and globally fit.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 April; accepted 26 October 2007.

Published online 18 November 2007.

- Austin, R. H., Beeson, K. W., Eisenstein, L., Frauenfelder, H. & Gunsalus, I. C. Dynamics of ligand binding to myoglobin. *Biochemistry* **14**, 5355–5373 (1975).
- McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267**, 585–590 (1977).
- Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **254**, 1598–1603 (1991).
- Frauenfelder, H., McMahon, B. H. & Fenimore, P. W. Myoglobin: the hydrogen atom of biology and a paradigm of complexity. *Proc. Natl Acad. Sci. USA* **100**, 8615–8617 (2003).
- Levinthal, C. Are there pathways for protein folding. *J. Chim. Phys. Phys.-Chim. Biol.* **65**, 44–45 (1968).
- Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167–195 (1995).
- Dobson, C. M., Sali, A. & Karplus, M. Protein folding: a perspective from theory and experiment. *Angew. Chem. Int. Ed.* **37**, 868–893 (1998).
- Williams, J. C. & McDermott, A. E. Dynamics of the flexible loop of triosephosphate isomerase: the loop motion is not ligand gated. *Biochemistry* **34**, 8309–8319 (1995).
- Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313**, 1638–1642 (2006).
- Palmer, A. G. NMR characterization of the dynamics of biomacromolecules. *Chem. Rev.* **104**, 3623–3640 (2004).
- Cui, Q. & Karplus, M. Catalysis and specificity in enzymes: a study of triosephosphate isomerase and comparison with methyl glyoxal synthase. *Adv. Protein Chem.* **66**, 315–372 (2003).
- Eisenmesser, E. Z. *et al.* Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117–121 (2005).
- Wolf-Watz, M. *et al.* Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nature Struct. Mol. Biol.* **11**, 945–949 (2004).
- Blanchard, S. C., Gonzalez, R. L., Kim, H. D., Chu, S. & Puglisi, J. D. tRNA selection and kinetic proofreading in translation. *Nature Struct. Mol. Biol.* **11**, 1008–1014 (2004).
- Ha, T. *et al.* Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism. *Proc. Natl Acad. Sci. USA* **96**, 893–898 (1999).
- Myong, S., Stevens, B. C. & Ha, T. Bridging conformational dynamics and function using single-molecule spectroscopy. *Structure* **14**, 633–643 (2006).
- Rothwell, P. J. *et al.* Multiparameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 reverse transcriptase: primer/template complexes. *Proc. Natl Acad. Sci. USA* **100**, 1655–1660 (2003).
- Schuler, B., Lipman, E. A. & Eaton, W. A. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* **419**, 743–747 (2002).
- Zhang, Z., Rajagopalan, P. T. R., Selzer, T., Benkovic, S. J. & Hammes, G. G. Single-molecule and transient kinetics investigation of the interaction of dihydrofolate reductase with NADPH and dihydrofolate. *Proc. Natl Acad. Sci. USA* **101**, 2764–2769 (2004).
- Vonrhein, C., Schlauderer, G. J. & Schulz, G. E. Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases. *Structure* **3**, 483–490 (1995).
- Müller, C. W., Schlauderer, G. J., Reinstein, J. & Schulz, G. E. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure* **4**, 147–156 (1996).
- Henzler-Wildman, K. A. *et al.* A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* doi:10.1038/nature06407 (this issue).
- Müller, C. W. & Schulz, G. E. Structure of the complex between adenylate kinase from *Escherichia coli* and the inhibitor Ap5A refined at 1.9 Å resolution. A model for a catalytic transition state. *J. Mol. Biol.* **224**, 159–177 (1992).
- Błaszczak, J., Li, Y., Yan, H. G. & Ji, X. H. Crystal structure of unligated guanylate kinase from yeast reveals GMP-induced conformational changes. *J. Mol. Biol.* **307**, 247–257 (2001).
- Faber, H. R. & Matthews, B. W. A mutant T4 lysozyme displays 5 different crystal conformations. *Nature* **348**, 263–266 (1990).
- Gardberg, A., Shuvalova, L., Monnerjahn, C., Konrad, M. & Lavie, A. Structural basis for the dual thymidine and thymidylate kinase activity of herpes thymidine kinases. *Structure* **11**, 1265–1277 (2003).
- Odintsov, S. G., Sabala, I., Bourenkov, G., Rybin, V. & Bochtler, M. Substrate access to the active sites in aminopeptidase T, a representative of a new metallopeptidase clan. *J. Mol. Biol.* **354**, 403–412 (2005).
- Gerstein, M., Schulz, G. & Chothia, C. Domain closure in adenylate kinase — joints on either side of 2 helices close like neighboring fingers. *J. Mol. Biol.* **229**, 494–501 (1993).
- Maragakis, P. & Karplus, M. Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase. *J. Mol. Biol.* **352**, 807–822 (2005).
- Karplus, M. & Kushick, J. N. Method for estimating the configurational entropy of macromolecules. *Macromolecules* **14**, 325–332 (1981).
- Ma, J. & Karplus, M. Ligand-induced conformational changes in ras p21: a normal mode and energy minimization analysis. *J. Mol. Biol.* **274**, 114–131 (1997).
- Bahar, I. & Rader, A. J. Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* **15**, 586–592 (2005).
- Miyashita, O., Onuchic, J. N. & Wolynes, P. G. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc. Natl Acad. Sci. USA* **100**, 12570–12575 (2003).
- Lou, H. & Cukier, R. I. Molecular dynamics of apo-adenylate kinase: a principal component analysis. *J. Phys. Chem. B* **110**, 12796–12808 (2006).
- Ha, T. *et al.* Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl Acad. Sci. USA* **93**, 6264–6268 (1996).
- Rhoads, D. G. & Lowenstein, J. M. Initial velocity and equilibrium kinetics of myokinase. *J. Biol. Chem.* **243**, 3963–3972 (1968).
- Nie, S., Chiu, D. T. & Zare, R. N. Probing individual molecules with confocal fluorescence microscopy. *Science* **266**, 1018–1021 (1994).
- Margittai, M. *et al.* Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1. *Proc. Natl Acad. Sci. USA* **100**, 15516–15521 (2003).
- Iwahara, J., Tang, C. & Clore, G. M. Practical aspects of ¹H transverse paramagnetic relaxation enhancement measurements on macromolecules. *J. Magn. Reson.* **184**, 185–195 (2007).
- Bruice, T. C. Computational approaches: reaction trajectories, structures, and atomic motions. Enzyme reactions and proficiency. *Chem. Rev.* **106**, 3119–3139 (2006).
- Hammes-Schiffer, S. & Benkovic, S. J. Relating protein motion to catalysis. *Annu. Rev. Biochem.* **75**, 519–541 (2006).
- Nagel, Z. D. & Klinman, J. P. Tunneling and dynamics in enzymatic hydride transfer. *Chem. Rev.* **106**, 3095–3118 (2006).
- Vendruscolo, M. & Dobson, C. M. Dynamic visions of enzymatic reactions. *Science* **313**, 1586–1587 (2006).
- Karplus, M. & McCammon, J. A. The internal dynamics of globular-proteins. *Crit. Rev. Biochem.* **9**, 293–349 (1981).
- Kuriyan, J. & Weiss, W. I. Rigid protein motion as a model for crystallographic temperature factors. *Proc. Natl Acad. Sci. USA* **88**, 2773–2777 (1991).
- Loria, J. P., Rance, M. & Palmer, A. G. A. TROSY CPMG sequence for characterizing chemical exchange in large proteins. *J. Biomol. NMR* **15**, 151–155 (1999).
- Mulder, F. A. A., Mittermaier, A., Hon, B., Dahlquist, F. W. & Kay, L. E. Studying excited states of proteins by NMR spectroscopy. *Nature Struct. Biol.* **8**, 932–935 (2001).
- Brooks, B. R. *et al.* CHARMM — a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank V. Orekhov at the Swedish NMR Centre for 800 MHz NMR spectrometer time, L. Kay for providing pulse programs, D. Korzhnev for sharing software for NMR relaxation data analysis and J. Hohlbein for Monte Carlo simulation software. We are grateful to K. O. Stetter for providing DNA isolated from *A. aeolicus* and the Advanced Biomedical Computing Center for CPU hours. This work was supported by NIH grants to D.K. and K.A.H.-W., a DOE grant to D.K., a fellowship from the American Heart Association to M.L., a Volkswagen Foundation grant to C.G.H. and M.O., and the Studienstiftung des Deutschen Volkes to M.O. The research at Harvard was supported in part by a grant from NIH to M.K.

Author Contributions K.A.H.-W., V.T., M.L. and M.O. contributed equally to this work. V.T. solved the X-ray structures with assistance from T.F., E.P., M.A.W. and G.A.P. NMR experiments were performed by K.A.H.-W., V.T. and M.W.-W., with assistance from D.K. Computation was carried out by M.L., with assistance and supervision from M.K. and D.K. FRET experiments were designed and performed by M.O. and K.A.H.-W., with assistance and supervision from C.G.H. and D.K. K.A.H.-W. and D.K. wrote the manuscript, and D.K. supervised all aspects of this work.

Author Information The x-ray structures of *Aquifex* Adk and Zn•Ap5A-bound *Aquifex* Adk are deposited in the Protein Data Bank under ID code 2RH5 and 2RGX, respectively. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.K. (dkern@brandeis.edu) or C.G.H. (huebner@physik.uni-luebeck.de).

METHODS

Protein expression, purification and preparation. $^2\text{H}/^{15}\text{N}$ -labelled Adk with protonated amides was prepared by dialyzing purified protein against 6 M urea, pH 7, for 24 h followed by refolding in NMR buffer (40 mM MOPS, pH 7.0, and 50 mM NaCl). Fluorescently labelled Adk was prepared by reacting 100 μM aliquots of Y52C/V145C-6 \times His-tag mutants with 15-fold molar excess of Alexa633-maleimide and Alexa488-maleimide. Excess dye was removed using two consecutive PD-10 columns pre-equilibrated with FRET buffer (20 mM TRIS, pH 7.0, 50 mM NaCl). The final dye and protein concentrations were determined by absorbance and Lowry assay, respectively. Excess MTSL was removed from spin-labelled Adk using two consecutive PD-10 columns pre-equilibrated with NMR buffer to yield final samples of 0.7–1.0 mM protein in NMR buffer. The diamagnetic form of the spin label was prepared by adding threefold excess ascorbate from a 1 M stock in NMR buffer, and the sample pH was re-adjusted to 7.0.

X-ray crystallography structure determination. For crystallization, selenomethionine-labelled protein was expressed in B834 (DE3) methionine auxotrophe cell line (Novagen) as described (<http://alfi.mrc-lmb.cam.ac.uk/~ramak/madms/segrowth.html>), with the addition of 1 mM DTT to all buffers. Crystals of apo Adk were grown by the sitting-drop method using purified enzyme at 27 mg ml $^{-1}$ in 10 mM HEPES, pH 7.5. The drop contained a 1:1 mixture of protein solution and mother liquor (28–30% (w/v) PEG 3,000, 200 mM sodium acetate, and 100 mM TRIS, pH 8.5). Initial epitaxially twinned crystals were used for streak-seeding, yielding single and epitaxially twinned crystals. Crystals of Adk complexed with Zn^{2+} •Ap5A were obtained by the same method in the presence of 1 mM Ap5A and 50 mM MgCl_2 . The mother liquor was 25% (v/v) PEG MME 550, 0.1 M MES, pH 6.5, 0.01 M ZnSO_4 .

Diffraction data were collected at room temperature from a single capillary-mounted crystal at an in-house Rigaku 300B rotating anode X-ray generator and at the synchrotron beamline X6A at Brookhaven National Laboratory. A three-wavelength multiple anomalous dispersion (MAD) data set was collected on a single selenomethionine-labelled Adk crystal at 100 K. Unliganded Adk crystals were cryo-protected using a 1:1 mixture of 70% (w/v) trehalose and were equilibrated with buffer. Co-complex crystals did not require any additional solutions for cryo-protection. Crystals were flash-cooled by immersion into liquid N_2 . Diffraction data were integrated and scaled with DENZO and SCALEPACK⁴⁹.

The co-complex structure of *Aquifex* Adk with Zn^{2+} •Ap5A was solved by molecular replacement using AMORE⁵⁰ in CCP4 (ref. 51) with 1AKE as the search model. A single copy of the molecule was located in the asymmetric unit (ASU). The model was manually rebuilt in O⁵² and refined with CNS⁵³. Model phases were used to calculate an anomalous difference map using FFT⁵⁴. Anomalous sites were modelled as zinc ions based on the presence of zinc sulphate in the mother liquor. This was confirmed from analysis of difference density maps. A final round of refinement was performed in REFMAC5 (ref. 55).

A combination of molecular replacement and MAD phasing techniques was required to determine the initial phases for the structure of ligand-free *Aquifex* Adk. Initial attempts at molecular replacement using AMORE and PHASER⁵⁶ with a number of adenylate kinase homologues located only a single copy in the asymmetric unit. A three-wavelength MAD data set was collected from a single crystal of selenomethionine-substituted Adk at wavelengths chosen to correspond to the peak, inflection point and high energy side of the selenium K edge as determined from an X-ray fluorescence scan of the Adk crystals (Supplementary Table 1). Determination of the location of the selenium sites using several methods yielded poor maps. The low quality of the maps is probably due to a low number of ordered selenomethionine sites in the asymmetric unit (6 methionines, excluding the 3 amino-terminal methionines, for 618 residues in the ASU), which resulted in phases of poor quality (0.68 and 0.53 figure of merit from SOLVE⁵⁷ and RESOLVE⁵⁸, respectively).

To improve the interpretability of the electron density maps and to ensure a consistent choice of origin, the initial molecular replacement solution was used to generate phases for the calculation of an anomalous difference Fourier map to locate nine selenium sites that were then used for MAD phasing. A new electron density map, calculated using combined molecular replacement and MAD phases, was interpretable and allowed us to locate the second molecule of Adk in the ASU. The phases calculated from the two-monomer model were combined with the MAD phases using sigmaA⁵⁹, and were improved using density modification in RESOLVE⁵⁸. An electron density map of satisfactory clarity was calculated with these density-modified phases and was used to guide manual rebuilding of the model in O. The model was refined with CNS and extended using automated structure building in ARP/wARP⁶⁰; electron density improvement was performed using the *Mycobacterium tuberculosis* (TB) structural genomics consortium bias removal server (<http://tuna.tamu.edu/>)⁶¹. After a number of rebuilding cycles, new fragments of density appeared outside of the

two-monomer model. ARP/wARP automatically built a poly-alanine model containing three α -helical fragments into this density.

Superposition of the complete monomer model over these fragments revealed a single docking possibility consistent with crystal packing. The third monomer was manually docked, rebuilt in O and refined with CNS. The final three-monomer model was refined to a final $R/R_{\text{free}} = 28/32$. Attempts to improve the R factors by refinement in the space group C222₁ and lower symmetry space groups C2, P2₁ and P1 did not affect R/R_{free} . The model was subsequently refined against the lower resolution, room temperature data. The final cycles of model building and refinement, including TLS refinement^{45,62}, were performed with Coot⁶³ and REFMAC5 (ref. 55). After TLS refinement, the final values of $R/R_{\text{free}} = 19.7/25.5$.

Three TLS models were tested and compared to the results of restrained refinement without inclusion of TLS parameters (Supplementary Table 2). The first TLS model treated all three molecules in the ASU as a single rigid group. The second treated each of the three monomers as separate rigid TLS groups. The third separated each molecule into the three standard nucleoside monophosphate domains: core (residues 1–29, 72–110 and 170–202), ATP lid (residues 111–169) and AMP lid (residues 30–71), giving a total of nine TLS groups for the ASU. TLS refinement was performed by first fixing the atomic B factors to 20 \AA^2 followed by ten cycles of TLS refinement and ten cycles of maximum likelihood restrained refinement of coordinates and B factors in REFMAC5. For all TLS models, all protein atoms were included for TLS refinement. Treatment of the ASU as a single TLS group (model 1) did not show any improvement over the isotropic B factor model (see Supplementary Table 2). However, both TLS models 2 and 3 led to improvements in both R and R_{free} when compared to the isotropic B-factor model. Model 3, which includes a TLS treatment for each domain of the three molecules in the ASU, provides the best fit to the data because it produces the lowest values for both R and R_{free} .

Crystal packing: conformational substates and B factors. The crystal environment of a protein is known to influence its conformation, particularly if the regions involved in crystal packing are flexible. Conformational heterogeneity of the mobile nucleotide-binding lids and less-ordered regions found within the three molecules in the ASU of apo Adk can be rationalized by considering the differences in crystal contacts for each molecule. Of the three molecules, molecule A has the most crystal contacts. Approximately 60% of the surface area of molecule A is buried in surface contact as compared to 39% and 40% of molecule B and molecule C, respectively (Supplementary Table 3 and Supplementary Figs 1 and 2). The significant amount of crystal contact and the uniform distribution of these contacts across molecule A probably limit its mobility and displacement within the ASU (Supplementary Fig. 1a), as judged by the quality of the electron density and by lower overall B factors and smaller TLS parameters (Supplementary Fig. 3 and Supplementary Table 2). In contrast, the core of molecule B has the poorest density as well as the highest overall B factors and largest TLS parameters compared to the core in the other two molecules (Supplementary Fig. 1b and Supplementary Table 2). This correlates well with the lack of crystal contacts in this region of molecule B (Supplementary Figs 1b and 2b). With the current data, we are unable to differentiate whether this disorder is caused by dynamic domain movement or static disorder of this region in the crystal.

Although molecules A and B have significant crystal contacts on the front and back side of their nucleotide-binding domains, where front and back refers to the surface of the domain facing towards or away from the nucleotide binding pocket, respectively, molecule C is missing contacts along the front side (Supplementary Figs 1c and 2c). The lack of crystal contact in the front regions of the nucleotide lids of molecule C and the presence of crystal contacts along the back of the lids are probably responsible for the observed partially closed conformation. The nucleotide lids of molecule A and molecule B are prevented from such compression by crystal-packing interactions with other symmetry copies (Supplementary Fig. 2a, b). The front side of molecule A interacts with a symmetry-related copy of itself and molecule B. These interactions limit the flexibility and displacement of the nucleotide lids. The ATP lid of molecule B, which is packed against the ATP lid and core of molecule A, is prevented from further displacement owing to the fact that the back of the ATP lid from molecule A is packed against the back of the AMP lid from molecule C. Having both domains pack into the interface between the nucleotide-binding lids of molecule B limits the movement of the nucleotide lids of molecule B. Molecule C, which is the most closed, only interacts with a portion of the AMP lid of molecule B. Because there are no interactions on the front side of the nucleotide lids, but there are packing interactions on the back side of the lid, the nucleotide lids are compressed towards the closed state.

In addition to the displacements leading to the three distinct substates in the crystal, different overall B factors are observed for molecule A, B and C. Whereas simulations of the three molecules in isolated water droplets show the largest

amplitude thermal fluctuations for the nucleotide lids with almost identical B factors for all three substates (Supplementary Fig. 3), crystallographic B factors do not agree with these computational results. Overall B factors differ significantly in the three molecules (see text above and Supplementary Fig. 3), and can be rationalized well by the differences in crystal contacts described above. To test that protein rigid body displacement contributes to the observed high overall B factors, we extracted the TLS component and the residual component from the overall B factors using TLSANL⁶⁴ (Supplementary Fig. 3). The results clearly show that rigid body displacements with the domains defined above (TLS refinement) account for the differences in the overall B factors. After TLS refinement, the residual B factors (excluding the contribution of the refined TLS parameters) of all residues in all three molecules are between 40 and 60 (red line in Supplementary Fig. 3). **Relaxation dispersion NMR spectroscopy.** Spectra were acquired on 800, 600 and 500 MHz Varian NMR spectrometers. Backbone assignments at different temperatures and for the Mg^{2+} •Ap5A-bound form were obtained by temperature titrations to transfer previously published backbone assignments¹³, and were confirmed by standard triple-resonance experiments. Relaxation data were fit to a two-site model describing the opening and closing of the nucleotide lids using the full Carver–Richards equation⁶⁵ with error analysis as described¹²:

$$R_2^{\text{eff}} = \frac{1}{2} \left(R_{2A}^0 + R_{2B}^0 + k_{\text{ex}} - \frac{1}{\tau_{\text{CP}}} \cosh^{-1} (D_+ \cosh \eta_+ - D_- \cosh \eta_-) \right) \quad (1)$$

$$\text{where} \quad D_{\pm} = \frac{1}{2} \left(\frac{\psi + 2\Delta\omega_N^2}{\sqrt{\psi^2 + \zeta^2}} \pm 1 \right), \quad \eta_{\pm} = \frac{\tau_{\text{CP}}}{\sqrt{2}} \sqrt{\psi^2 + \zeta^2} \pm \psi,$$

$$\psi = (R_{2A}^0 - R_{2B}^0 - p_A k_{\text{ex}} + p_B k_{\text{ex}})^2 - \Delta\omega_N^2 + 4p_A p_B k_{\text{ex}}^2,$$

$$\zeta = 2\Delta\omega_N (R_{2A}^0 - R_{2B}^0 - p_A k_{\text{ex}} + p_B k_{\text{ex}}) \text{ and } v_{\text{CPMG}} = \frac{1}{2\tau_{\text{CP}}}. \text{ Here, } k_{\text{ex}} \text{ is the rate of}$$

exchange between states A and B, R_{2A}^0 and R_{2B}^0 are the inherent transverse relaxation rates in state A and B, respectively, and are assumed to be the same, p_A and p_B are the populations of state A and B, respectively, $\Delta\omega_N$ is the chemical shift difference between the two exchanging states, and τ_{CP} is the time between 180° pulses in the NMR experiment. We note that a two-state exchange is certainly an oversimplified model because both lids can move. However, the good correlation between $\Delta\omega$ calculated from the CPMG experiments and the $\Delta\omega$ measured for the open and closed conformations¹³ (Supplementary Table 4 and Supplementary Fig. 4) justifies such a simplification:

For Mg^{2+} •Ap5A-bound Adk, $\Delta\omega_N$ values from the CPMG relaxation dispersion experiments ($\Delta\omega_N^{\text{CPMG}}$) were compared to the $\Delta\omega_N$ values between the open and closed state ($\Delta\omega_N^{\text{HSQC}}$) measured directly from the ^{15}N – ^1H heteronuclear single-quantum coherence (HSQC) spectra of apo (open) and Mg^{2+} •Ap5A-saturated (closed) Adk samples (Supplementary Table 4). Although both Mg^{2+} •Ap5A-saturated Adk and apo Adk sample open and closed conformations, as measured in the dispersion experiments, the HSQC spectra are indicative of the chemical shifts of the closed and open state, respectively, because the populations are highly skewed. For a number of residues spread throughout the protein but remote from the active site, very good agreement between $\Delta\omega_N^{\text{CPMG}}$ and $\Delta\omega_N^{\text{HSQC}}$ is observed. All of the residues with discrepancies between $\Delta\omega_N^{\text{CPMG}}$ and $\Delta\omega_N^{\text{HSQC}}$ can be explained by direct proximity to the inhibitor or proximity to additional metal-binding sites (Supplementary Fig. 4 and Supplementary Table 4). In the crystal structure of Adk bound to Zn^{2+} •Ap5A, several Zn^{2+} -binding sites are observed on the surface of the protein in addition to the Zn^{2+} in the active site. The side chains of residues that chelate the metal in these additional binding sites are shown in orange in Supplementary Fig. 4. The replacement of Mg^{2+} by Zn^{2+} in the crystal is due to the mother liquor. The NMR experiments are performed under saturating concentrations of Mg^{2+} •Ap5A. The large NMR chemical shift differences between apo Adk and Mg^{2+} •Ap5A-saturated Adk close to these Zn^{2+} sites is probably caused by Mg^{2+} binding to these locations.

Paramagnetic relaxation enhancement NMR. DTNB (5,5'-dithiobis(2-nitrobenzoic acid)) assays were performed⁶⁶ before and after labelling to confirm complete reaction of the Cys residue with the spin label. The HSQC spectra of the paramagnetic and diamagnetic forms of MTSL-*Aquifex* Adk(Y52C) were processed with only 5–10 Hz exponential broadening because a lorentzian line-shape is necessary for the single relaxation time analysis of paramagnetic relaxation effects⁶⁷. Distances were calculated from the intensity ratio of the HSQC spectra using R_2 values measured on the diamagnetic sample⁶⁸ and the following equations^{39,67,69}:

$$\frac{I_{\text{para}}}{I_{\text{dia}}} \approx \frac{R_2 \exp(-R_2^* t)}{R_2 + R_2^*} \quad (2)$$

$$d = \left[\frac{K}{R_2^*} \left(4\tau_c + \frac{3\tau_c}{1 + \omega_H^2 \tau_c^2} \right) \right]^{1/6} \quad (3)$$

where I_{para} and I_{dia} are the peak heights in the HSQC spectra of the paramagnetic and diamagnetic samples, respectively, R_2 is the amide proton R_2 measured on the diamagnetic sample, R_2^* is the additional relaxation due to the spin label, ω_H is the proton larmor frequency, K is the constant $1.23 \times 10^{-32} \text{ cm}^6 \text{ s}^{-2}$ for a nitroxide radical, and d is the distance between the proton and paramagnetic electron. τ_c is the electron–amide proton correlation time, which was assumed to be equal to the global tumbling time of the unlabelled protein²²—a good approximation in the case of a nitroxide spin label^{39,67}. The calculated distances from the PRE experiment on ligand-free Adk closely resemble the distances for the closed state except a few residues in the C-terminal helix have calculated distances a little shorter than in the closed state (Fig. 3e). This is probably due to the flexibility of the C-terminal helix as evidenced by R_{ex} for a number of residues in this helix (Fig. 3c).

NMR experiments on FRET-labelled samples. To compare directly the NMR and single-molecule FRET dynamics, several controls using the FRET-labelled samples for NMR experiments were performed. We note that the experiments described here do not provide an exact quantitative comparison owing to the technical difficulties addressed below; however, they do yield the accuracy needed for the conclusions drawn in the manuscript.

$^{15}\text{N}/^1\text{H}$ -*Aquifex* Adk(V145C/Y52C) with Alexa633 and Alexa488 attached at the cysteine residues was produced for NMR experiments on FRET-labelled Adk. First, $^{15}\text{N}/^1\text{H}$ HSQC spectra were compared between the wild-type and FRET-labelled sample. Attachment of the bulky hydrophobic dyes resulted in severe line-broadening of all resonances, which was dependent on the protein concentration. This is indicative of transient intermolecular aggregation caused by the hydrophobic dyes. However, a decent NMR spectrum could be obtained at a protein concentration of 130 μM at 30 °C and 40 °C (Supplementary Fig. 5). Residues near the dye-attachment sites showed additional chemical shifts or extreme line broadening, as expected from the vicinity of the dyes and fluctuations of the bulky hydrophobic dyes resulting in intermediate timescale exchange. However, residues in the core of the protein could still be used to compare protein conformation between the wild-type and FRET-labelled Adk. Supplementary Fig. 5 shows that the chemical shifts of residues not near the fluorescent dyes are identical for wild-type and FRET-labelled apo Adk (blue and green spectra, respectively) and also for wild-type and FRET-labelled Mg^{2+} •Ap5A-bound Adk (black and red spectra, respectively). This indicates that the distribution of mostly open conformations in apo Adk is not altered by attachment of the fluorescent dyes, and that the dyes do not affect Mg^{2+} •Ap5A binding and lid closure.

Second, to compare quantitatively the single-molecule FRET kinetics (performed on protonated apo Adk at room temperature) with NMR kinetics, NMR dispersions were analysed on protonated apo Adk at 20 °C. ^{15}N CPMG relaxation dispersion NMR experiments were performed initially on $^{15}\text{N}/^1\text{H}$ apo Adk at 20 °C to compare the results with those previously obtained during turnover at 20 °C (ref. 13). However, the amplitude of exchange was small and the rate of exchange was fast (Supplementary Fig. 6a), preventing quantitative analysis of the data. Therefore, the experiments were repeated with $^{15}\text{N}/^2\text{H}$ apo Adk at 10 °C to increase the amplitude, decrease the rate and quantitatively determine the exchange parameters, as described in the main text. For a more quantitative comparison between the NMR and single-molecule FRET experiments, further analysis of the 20 °C protonated apo Adk relaxation dispersion profiles was performed to estimate the range of exchange rates compatible with the data. For this aim, residues with exchange in the $^{15}\text{N}/^1\text{H}$ apo Adk 20 °C data set were fitted with fixed exchange rates between 2,000 s^{-1} and 10,000 s^{-1} , and the R_2^0 values (R_2^{eff} at infinite v_{CPMG} with exchange fully suppressed) obtained from those fits were compared to the R_2^0 values of residues with no exchange (R_2^{eff} flat as a function of v_{CPMG}) (Supplementary Fig. 6b). As shown by the histogram of R_2^0 values across the protein, R_2^0 values fall between 15 s^{-1} and 23 s^{-1} (with a few exceptions). Comparison of this histogram with the distribution of fitted R_2^0 values for exchanging residues provides a means of estimating which exchange rates are most probable. Supplementary Fig. 6b shows that exchange rates between 2,000 s^{-1} and 7,000 s^{-1} are most consistent with the NMR data, in agreement with the single-molecule FRET experiments where the opening/closing rate at room temperature was measured as about 8,500 s^{-1} .

Third, the kinetics of lid motion was compared in wild-type and the FRET-labelled samples. The increased line-width of FRET-labelled Adk resulted in greater errors in the R_2 values, obscuring the small exchange contribution to the observed relaxation rate for apo Adk. Higher temperatures reduced the line-width but increased the rate out of the range accessible for CPMG experiments in the case of apo Adk. However, CPMG experiments could be performed on FRET-labelled Adk saturated with Mg^{2+} •Ap5A at 30 °C (Supplementary Fig. 7). The exchange rate of wild-type Mg^{2+} •Ap5A-bound Adk is about 300 s^{-1} at 20 °C (see main text). At 30 °C, the exchange rate for Mg^{2+} •Ap5A-bound FRET-labelled Adk is $600 \pm 300 \text{ s}^{-1}$ (mean \pm s.d.) (Supplementary Fig. 7). The error

is large, as expected because of slower tumbling, low protein concentration and fewer residues that can be monitored. The rate is faster than at 20 °C, as expected, and is comparable to the single-molecule FRET experiments performed at room temperature, which found an exchange rate of about 390 s⁻¹ for lid opening/closing.

Simulations. Molecule A, B and C of the apo *Aquifex* Adk crystal structure were simulated separately in water spheres with the stochastic boundary potential method^{70,71}, which prevents the solvent molecules from escaping from the spheres and provides a heat bath mimicking the surrounding solvent. The water molecules were represented as modified TIP3P⁷². The program Solvate 1.0 (by H. Grubmüller) was used to add a thick layer of water molecules and counter ions around the crystal structure of each molecule. Solvent molecules beyond 45 Å from the centre of each protein were deleted, leaving the minimum distance between the surfaces of water spheres and any protein atoms to be about 16 Å. The CHARMM 22-protein all-atom force field⁷³ was used in the simulations with the CMAP backbone energy correction included⁷⁴. Non-bonded interactions beyond 14 Å were shifted to be zero. The bond lengths between hydrogen and heavy atoms were fixed by the SHAKE algorithm⁷⁵. The time-step was 2 fs. To obtain the systems for starting the simulations, the protein and their surrounding solvent molecules were first minimized with gradually reducing harmonic constraints on protein backbone atoms. They were then slowly heated in 30 steps from 0 K to 300 K with 4 ps equilibration at each temperature step. The systems were then equilibrated at 300 K for 40 ps. The production runs were at 300 K for 10 ns. For the normal mode analysis³¹, the EEF1 implicit solvent model⁷⁶ was used. Quasi-harmonic (also called principal component) analysis³⁰ of the molecular dynamics trajectory was performed by diagonalizing the equal time mass weighted atom fluctuation cross-correlation matrix.

To investigate whether the ATP and AMP lids open and close in a correlated or independent fashion, we computed the normalized fluctuation correlation σ_{ij} between residual pairs as⁷⁷

$$\sigma_{ij} = \frac{\langle (\vec{r}_i - \langle \vec{r}_i \rangle) \bullet (\vec{r}_j - \langle \vec{r}_j \rangle) \rangle}{\sqrt{\langle (\vec{r}_i - \langle \vec{r}_i \rangle)^2 \rangle} \sqrt{\langle (\vec{r}_j - \langle \vec{r}_j \rangle)^2 \rangle}} \quad (4)$$

in which \vec{r}_i (\vec{r}_j) is the coordinate of the *i*th (*j*th) residue at a snapshot while the bracket indicates the average over all snapshots: σ_{ij} normalized to be between -1 and 1. When the motion between two residues is mostly anti-correlated (that is, when two residues are moving in opposite directions), σ_{ij} is negative and is coloured blue in Supplementary Fig. 8a. When the motion between two residues is mostly correlated (that is, when the two residues are moving in the same direction), σ_{ij} is positive and is coloured red in Supplementary Fig. 8a. When the motion of two residues is almost independent of each other, σ_{ij} is close to 0 and is coloured white in Supplementary Fig. 8a.

It is worth noting that σ_{ij} is influenced strongly by how the simulation trajectory is aligned onto the reference frame (Supplementary Fig. 8). When the simulation trajectory is aligned onto the crystal structure of the closed state by superimposing the whole protein, ATP and AMP lids seem to have anti-correlated motion, as indicated by the large blue block in the lower half-triangle of Supplementary Fig. 8a. In contrast, if only the core domain is superimposed during the alignment, the motion of the ATP lid is shown to be mostly independent of the motion of the AMP lid, as indicated by the large white block in the upper-half triangle of Supplementary Fig. 8a. Because our experimental NMR chemical shift data has shown that there are no internal conformational transitions in the core domain, the core domain serves as a better reference structure than the whole protein to study how the ATP and AMP lids fluctuate. Therefore, the core domain was used as a reference in the other analyses in this paper. Here, the correlation between the motion of the two lids is 0.12. The lack of correlation is further illustrated in a plot of the root mean square deviation fluctuations of the AMP and the ATP lid relative to the closed conformation (Supplementary Fig. 8b).

To investigate whether the simulation trajectories starting from molecule A, B and C of the apo crystal structure overlap with each other, we computed the backbone r.m.s. deviation between the closed state and snapshots of the simulations starting from the crystal structures of molecule A, B and C (Supplementary Fig. 9). Three conclusions can be drawn from the results. First, essentially the same conformational space is sampled in 10 ns molecular dynamics simulations starting from molecule A, B or C. Second, the fluctuations are centred around substate A. Third, the energy barriers between the crystallographically detected substates A, B and C are low because the transitions between them are easily sampled in nanosecond timescale simulations.

Single-molecule spectroscopy. Labelling sites in the lids were selected on the basis of distance changes between the open and closed conformations, considering only non-charged surface-exposed residues that faced away from the active site. Alexa488 and Alexa633 (Molecular Probes) were used as a FRET pair owing

to their high photostability and quantum yield and because of the well-suited Förster radius R_0 of approximately 48 Å. The reactivity of the cysteines was checked using DTNB⁶⁶ before labelling with the dyes. The labelling efficiency was between 98% and 100% as determined by a DTNB assay (which detects non-reacted Cys residues) and by absorbance. The protein concentration of the stock solution was determined by absorbance at 280 nm before labelling. After the reaction with the dyes, remaining free dye was removed by two PD-10 columns. No free dye could be detected after the second column. The concentration of the dyes attached to the protein was determined by absorbance, yielding 99–100% labelling efficiency. The ratio of the two dyes on the protein was 0.9:1.

The fluorescence quantum yields of the dyes bound to the protein were determined relative to the quantum efficiencies of the free dyes provided by the manufacturer. Single-site Cys mutants (Y52C, V145C) were made and reacted with either Alexa488 or Alexa633 to measure site-specific quantum yields. We determined quantum efficiencies for the donor Alexa488 of $Y_D^{ATP} = 0.77$ when attached to the ATP lid and of $Y_D^{AMP} = 0.74$ when attached to the AMP lid. For the acceptor Alexa633, those numbers are $Y_A^{ATP} = 0.53$ and $Y_A^{AMP} = 0.55$, respectively. The Förster radius R_0 was calculated according to the equation⁷⁸:

$$R_0 = (8.79 \times 10^{-5} n^4 Y_D J \kappa^2)^{(1/6)} (\text{Å}) \quad (5)$$

where n is the refractive index of the medium, Y_D is the donor fluorescence quantum yield, κ^2 is the orientation factor, and J is the overlap integral of donor emission and acceptor absorption. The latter was computed from the emission spectrum of the protein-bound donor and the absorption spectrum of the protein-bound acceptor, respectively. The site-dependent subtle changes of Y_D have negligible influence on R_0 owing to the inverse sixth power dependence. Time-resolved fluorescence polarization experiments yielded steady-state anisotropy values smaller than 0.1 for both dyes, irrespective of the labelling site. Therefore, the dyes are freely rotating, justifying $\kappa^2 = 2/3$ for the calculation of R_0 .

Samples with labelled molecules were diluted to a concentration of 100 pM in FRET buffer. Tween-20 (0.01% v/v, Carl Roth) helped to prevent surface adhesion of the molecules. For measurements of the molecules in presence of the inhibitor, 4 mM Ap5A and 2 mM MgCl₂ were added.

For the dye Alexa633, dark states on two different timescales—a 100 μs timescale and rare dark states lasting considerably longer (up to about 100 ms)—are observed in the experiments with immobilized protein (Fig. 5). We attribute the short dark states to triplet excursions and the longer ones to reversible photochemical reactions, probably a radical ionic state⁷⁹. Collective donor and acceptor dark states on the microsecond timescale are indicative of efficient singlet-triplet annihilation^{80,81}. This means that both dyes do not emit light if the acceptor is in the triplet state, essentially because this triplet state acts as an energy trap (see the section about time-resolved experiments).

Confocal optical microscope. Single-molecule measurements were performed with a home-built sample-scanning optical confocal microscope. Excitation light was provided by two lasers, a continuous wave external cavity diode laser operating at 488 nm (Protera, Novalux) and a pulsed diode laser delivering 100-ps pulses at 635 nm (LDH-P-635 + PDL 808 driver, Picoquant). Both lasers were fed into a single-mode fibre for spatial filtering. The light from the fibre was collimated by a ×4 microscope objective and reflected off a multichroic beam-splitter (z405/488/633, Chroma). Excitation light was eventually focused by a microscope objective (Nikon CFI Plan Apochromat 60 × 1.25 WI for diffusion measurements and CFI Plan Apochromat 100 × 1.4 oil for immobilized measurements) into the sample solution or onto the surface of a glass slide. A piezo-driven stage (P-734, Physik-Instrumente) operated in closed-loop allowed for positioning of the focal spot with nanometre precision.

Fluorescent light collected by the microscope objective passed several filters before being focused onto the detectors: the multichroic beam-splitter, a long-pass filter suppressing the 488 nm laser (HQ500LP, Chroma), and an additional bandpass filter (HQ532/70, Chroma) for the donor and a longpass filter (HQ650LP, Chroma) for the acceptor. Emission was split by a dichroic mirror (595DCXR, Chroma) into donor and acceptor components. The active area of the single-photon counting modules (SPCM AQ14, Perkin-Elmer) served as the confocal pinhole. The overall fluorescence-detection efficiency was 0.16 for the donor and 0.13 for the acceptor. Pulses from the detectors were fed into a TCSPC board (TimeHarp200, Picoquant) operating in the time-tagged time-resolved mode. The whole microscope was controlled by a programmable digital-signal processing unit (ADWin Gold, Jäger) and home-written software (LabView).

Shot-noise-limited performance of the setup^{82,83} was tested with a sample of fluorescent-energy-transfer latex spheres (TetraSpeck 0.1 μm, Molecular Probes) immersed in water. We chose this sample for its homogeneity and for the absence of transfer efficiency fluctuations. A FRET efficiency histogram of the reference sample along with a Monte Carlo simulation is shown

in Supplementary Fig. 11. The excitation intensity for this experiment was chosen in such a way that the detected fluorescence rate matched the rate in the Adk experiments. In the simulation, freely diffusing particles with poissonian emission in two channels with a distribution corresponding to the experiment were modelled. The almost perfect match of the experimental with the simulated histogram is clear evidence for shot-noise-limited performance of the microscope.

Single-molecule FRET diffusion measurements. For burst experiments, the donor was excited continuously at 488 nm with 40 μ W, whereas the acceptor was excited directly at 635 nm with 10 μ W by picosecond pulses at a repetition rate of 10 MHz. Detected photons were sorted by their arrival time (time resolution of about 35 ps) with respect to the pulsed laser trigger (10 MHz) for assignment to the respective excitation sources using a time window of 20 ns after the pulse, accounting for the mean lifetime of the acceptor of 3.1 ± 0.2 ns. Counted photons of each detector channel were binned at 1 ms and corrected for background counts, quantum yields of fluorescence of the dyes, and quantum efficiency of detection (including collection efficiency, filter transmission and detector efficiency). Fluorescence bursts were identified using two threshold criteria—one threshold for the integrated emission after donor excitation and one for emission on direct excitation of the acceptor—to select for molecules bearing both a donor and acceptor. Partial averaging owing to conformational transitions during transit through the focus was taken into consideration in the calculation of distances from E_i histograms. Because a direct calculation of absolute distances from FRET efficiencies is hampered by the difficulty of determining directly photon detection efficiencies in both detection channels and by the potentially biasing effect of dyes that are not freely rotating¹⁸, the Mg^{2+} •Ap5A-bound form was used as a reference point for the E_i value corresponding to the fully closed state.

Single-molecule FRET on immobilized samples. Glass slides were derivatized with *N*-((3-trimethoxysilyl)propyl)ethylene-diamine triacetic acid (Fluorochem) according to the manufacturer's procedure, and were chelated with Ni^{2+} (ref. 84). First, confocal images were taken at 100 nW continuous excitation at 488 nm to find molecules showing energy transfer, ensuring that both donor and acceptor were present. Individual molecules were then positioned in the focus and were excited continuously at 488 nm with a laser power of 2 μ W for Mg^{2+} •Ap5A-bound Adk and of 10 μ W and 30 μ W for ligand-free Adk.

The entire time traces of the two molecules shown in Supplementary Fig. 16a and Fig. 5d, but with an integration time of 10 ms shown in Supplementary Fig. 14a, b, correspondingly demonstrate the digital bleaching behaviour of the dyes characteristic of single-molecule observations. The temporal resolution is dictated essentially by the intensity of the signal. The laser intensity cannot be increased to arbitrary levels, because of the imminent photo-destruction. As seen in Fig. 5, the applied laser intensities enabled detection of individual events. Of a total of about 200 time traces for the ligand-free Adk and for Mg^{2+} •Ap5A-bound Adk, only 25 survived long enough for a statistical state analysis.

Changes between the open and closed states were identified on the basis of an extended threshold criterion. Instead of a single threshold value, a range of E_i values—a transition zone—was defined where the gaussian distributions of the open and closed states overlap in excess of 0.3. This overlap is caused by the shot noise of photon detection and by unresolved conformational substates. Within this range, an unambiguous assignment to either state is not possible. Therefore, E_i crossing of the transition zone indicates a change of state. The time instance where E_i is leaving the transition zone is identified as the change-of-state time. The lifetimes determined from the time traces were fitted exponentially excluding the first data point in the histograms because it contains shot-noise-induced apparent transitions.

To exclude artificial photo-induced effects on the apparent FRET dynamics, we performed control experiments at a threefold-increased laser intensity of 30 μ W. As Supplementary Fig. 15 shows, the rates at 10 μ W and 30 μ W laser intensity agree very well. For comparison, we performed a sample correlation analysis for immobilized apo Adk molecules. Here, an excitation intensity of 30 μ W delivered the least-noisy correlation.

The specificity of His-tagged Adk immobilization was tested in the following way. First, functionalized coverslips preincubated with imidazole did not show any binding (Supplementary Fig. 12a, b). Second, repeated washing of the coverslips with tethered Adk did not remove any molecules (Supplementary Fig. 12c, d). Third, washing with imidazole completely removed bound Adk (Supplementary Fig. 12e).

We performed a luciferase assay on the immobilized protein to ensure its functionality (Supplementary Fig. 13). A drop of buffer containing firefly luciferase, ADP and Mg^{2+} was deposited on top of a coverslip with immobilized labelled Adk.

For the assay, three different surface densities between 120–350 molecules per μm^2 were chosen. The surface density was determined from the total laser-induced fluorescence intensity of the tethered labelled Adk. For calibration, the

intensity of an image with a surface density low enough to allow for molecule counting was used (data not shown). The surface densities correspond to concentrations if the molecules were dissolved in the drop of buffer above the coverslip in the range of 0.1 nM to 0.3 nM. To test that all Adk molecules are tethered to the coverslip and are not released into solution, fluorescence correlation spectroscopy (FCS) was performed in the buffer solution above the coverslip. The FCS curve essentially resembled that of the buffer, meaning that only the same weakly fluorescent contaminants are present, and was clearly distinguishable from a 0.1 nM Adk solution measured as a reference.

Supplementary Fig. 13a shows the increase of luciferase chemoluminescence as a function of time for three different Adk surface densities. The assay performed under identical conditions on the coverslip with known ATP concentrations was used for calibration (Supplementary Fig. 13b). From the linear fits in Supplementary Fig. 13a, a turnover rate of $42 \pm 10 \text{ s}^{-1}$ (mean \pm s.d.) per immobilized Adk molecule was obtained—a value very similar to the turnover rate for Adk measured with the coupled enzymatic assay. As the inset in Supplementary Fig. 13a demonstrates, there is a linear relationship between the Adk density and the slope of the chemoluminescence increase. This is proof for the quantitative validity of our surface-based assay.

Monte Carlo simulation of single-molecule time traces. To calculate the true opening and closing rates from the apparent rates, we generated simulated fluorescence intensity time traces (Supplementary Fig. 17). For the simulated time-traces, 10,000 exponentially distributed open and closed states including triplet dynamics were generated using Monte Carlo methods with a time resolution of 100 ns (identical to time resolution of photon counting). The data were then digitized in time bins according to the experimental conditions. Finally, the simulated time traces were analysed in exactly the same way as the experimental time traces. This procedure was iterated while changing the input parameters (lifetimes of open and closed states), until the analysis yielded the experimentally obtained apparent opening and closing rates (Supplementary Fig. 17a–d).

The simulations exemplify the following facts. First, they show clearly that the apparent rates extracted by a simple exponential fit of the experimental data (Supplementary Fig. 17a) yield slightly underestimated rates for ligand-free Adk owing to missed events (Supplementary Fig. 17c). Using simulations, the missed events can be predicted and then used to calculate the opening and closing rates. Our approach is similar to the missed event analysis for FRET data¹⁴ and for single-molecule ion channel recording⁸⁵, however, with triplet dynamics superimposed. Second, in contrast, for Mg^{2+} •Ap5A-bound Adk, the frequency of transitions is much slower than in ligand-free Adk (compare Supplementary Fig. 17a and 17b); this is specifically true for the opening rate, in agreement with our NMR data (Fig. 3a, b and Supplementary Fig. 6), resulting in a small overall number of transitions for a single time trace. This fact hampers an accurate states analysis. Moreover, the transition frequencies are now on the order of the occurrence of triplet dark states, significantly reducing the longer dwell times of the closed states (Supplementary Fig. 17b). Monte Carlo simulations further exemplify this fact, because the apparent opening rates (110 s^{-1}) are faster than the 'real' opening rates (40 s^{-1} , the input rates of the simulations) for Mg^{2+} •Ap5A-bound Adk. Third, the measured conformational transitions are clearly above the shot noise (Supplementary Fig. 17e, f), and our threshold analysis is robust for distinguishing between conformational transitions and shot-noise fluctuations (see Supplementary Fig. 17e, f). However, in the case of Mg^{2+} •Ap5A-bound Adk, shot noise does result in detectable reduction of the long dwell times of the closed state (Supplementary Fig. 17b, d, f). The acquisition of longer time traces for Mg^{2+} •Ap5A-bound Adk was not possible because of faster bleaching of the acceptor relative to ligand-free Adk owing to the higher energy transfer efficiency and consequently preferred excitation of the acceptor. Fourth, simulated and experimental time traces look similar in respect to 'noisiness', supporting the fact that the observed fluctuations in E_i within states are mainly caused by shot noise. Fifth, the E_i histograms of the simulated traces resemble those of the experimental traces quite nicely, whereas the corresponding histograms of pure shot noise are clearly distinct. There is some additional width in the histograms from the experiment, suggesting the possibility that there are more than just two states.

Correlation analysis of single-molecule time traces. Auto- and cross-correlation functions calculated for individual apo Adk molecules, and subsequently averaged, are shown in Fig. 5g. A global fit was performed with a model similar to that used in ref. 38 with included singlet–triplet annihilation:

For the acceptor and donor autocorrelation:

$$AC(t) = 1 - f(t) + A + B / ((1 - C)(1 - T))(1 - C + Ce^{-kt})(1 - T + Te^{-t/\tau_2 T}) \quad (6)$$

and for the cross-correlation:

$$CC(t) = 1 - f(t) + A + B(1 + Ce^{-kt})(1 + Te^{-t/\tau_1}) \quad (7)$$

were used as fitting functions, where A and B are constants, T is the triplet

fraction, τ_T is the triplet decay, and C is the additional correlated or anticorrelated fraction with k , the overall FRET dynamics rate constant. τ_T was determined by FCS of diffusing molecules and set to be 3.5 μ s, whereas the rate constant k was fit simultaneously. The function $f(t)$ was used to describe the drop of the correlation function owing to the cutoff effect.

In the donor–acceptor cross-correlation function, an anti-correlation is prominent for the times between 10^{-4} s and 10^{-3} s. Both autocorrelation functions, that of the donor and that of the acceptor, show a correlation with the same time constant. This finding of an anti-correlation in the cross-correlation function with corresponding decays in the correlation functions of the donor and the acceptor provides clear evidence for E_t fluctuations. The overall rate of these fluctuations is $7,000 \pm 1,500$ s $^{-1}$ (mean \pm s.d.), in very good agreement with the overall rate of $8,500 \pm 700$ s $^{-1}$ (mean \pm s.d.) obtained from state analysis. This rate also matches quite well the total rate obtained from NMR.

At shorter times (few microseconds), a correlation is present in all correlation functions, in the cross-correlation as well as in the auto-correlation functions. We assign this correlation time of 3.5 μ s to triplet excursions of the donor and the acceptor, respectively, apparently leading to correlated intensity fluctuations of both dyes. This correlation again strongly indicates efficient singlet–triplet annihilation. Overall triplet correlation times for the Adk-bound donor of 3.9 ± 0.2 μ s (mean \pm s.d.) and for the Adk-bound acceptor of 3.8 ± 0.2 μ s (mean \pm s.d.) were independently determined in solution for excitation rates comparable to those in the time-resolved FRET experiments (data not shown).

For inhibitor-bound Adk, we find similar behaviour (Supplementary Fig. 16d). However, the amplitude of the anti-correlated term in the cross-correlation function is extremely small because of the large ratio between the opening/closing rates expected from NMR dynamics. Therefore, a fit of the experimental data, such as for Apo Adk, yields very large errors. We can only give the order of magnitude of the overall rate of these fluctuations, which is 1,000 s $^{-1}$, in reasonable agreement with the state analysis and with the kinetics from NMR spectroscopy.

Single-label control experiments. An independent control experiment was performed on singly donor- or acceptor-labelled Adk to rule out that acceptor or donor dye fluctuations and not true energy transfer are causing the observed transfer efficiency fluctuations. The result supports our finding from the states analysis and the cross-correlation analysis because the donor- and acceptor-only time traces are shot-noise-limited with superimposed triplet dark states (Supplementary Fig. 18).

On request, we can provide all original time traces for those who want to perform their own types of analysis.

49. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillating mode. *Methods Enzymol.* **276**, 307–326 (1997).
50. Navaza, J. AMORE — an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
51. Collaborative Computational Project, Number 4. The CCP4 Suite — programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
52. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
53. Brunger, A. T. et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
54. Immirzi, A. in *Crystallographic Computing Techniques* (ed. Ahmed, F. R.) 399–412 (Munksgaard, Copenhagen, 1976).
55. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
56. Read, R. J. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr. D* **57**, 1373–1382 (2001).
57. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
58. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
59. Read, R. J. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42**, 140–149 (1986).
60. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463 (1999).
61. Reddy, V. et al. Effective electron-density map improvement and structure validation on a Linux multi-CPU web cluster: the TB structural genomics consortium bias removal Web service. *Acta Crystallogr. D* **59**, 2200–2210 (2003).
62. Schomaker, V. & Trueblood, K. N. On rigid-body motion of molecules in crystals. *Acta Crystallogr. B* **24**, 63–76 (1968).
63. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
64. Howlin, B. & Butler, S. A. Moss, D. S., Harris, G. W. & Driessen, H. P. C. TLSANL — TLS parameter-analysis program for segmented anisotropic refinement of macromolecular structures. *J. Appl. Cryst.* **26**, 622–624 (1993).
65. Carver, J. P. & Richards, R. E. A general two-site solution for the chemical exchange produced dependence of T2 upon the Carr–Purcell pulse separation. *J. Magn. Reson.* **6**, 89–105 (1972).
66. Riddles, P. W., Blakeley, R. L. & Zerner, B. Ellman's reagent: 5,5'-dithiobis(2-nitrobenzoic acid) — a reexamination. *Anal. Biochem.* **94**, 75–81 (1979).
67. Battiste, J. L. & Wagner, G. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* **39**, 5355–5365 (2000).
68. Donaldson, L. W. et al. Structural characterization of proteins with an attached ATCUN motif by paramagnetic relaxation enhancement NMR spectroscopy. *J. Am. Chem. Soc.* **123**, 9843–9847 (2001).
69. Solomon, I. & Bloembergen, N. Nuclear magnetic interactions in the HF molecule. *J. Chem. Phys.* **25**, 261–266 (1956).
70. Brooks, C. L. & Karplus, M. Deformable stochastic boundaries in molecular-dynamics. *J. Chem. Phys.* **79**, 6312–6325 (1983).
71. Brünger, A., Brooks, C. L. & Karplus, M. Stochastic boundary-conditions for molecular-dynamics simulations of ST2 water. *Chem. Phys. Lett.* **105**, 495–500 (1984).
72. Price, D. J. & Brooks, C. L. A modified TIP3P water potential for simulation with Ewald summation. *J. Chem. Phys.* **121**, 10096–10103 (2004).
73. MacKerell, A. D. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
74. MacKerell, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400–1415 (2004).
75. Ryckaert, J. P., Cicotti, G. & Berendsen, H. J. C. Numerical-integration of cartesian equations of motion of a system with constraints — molecular-dynamics of N-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
76. Lazaridis, T. & Karplus, M. Effective energy function for proteins in solution. *Proteins* **35**, 133–152 (1999).
77. Ichiye, T. & Karplus, M. Collective motions in proteins — a covariance analysis of atomic fluctuations in molecular-dynamics and normal mode simulations. *Proteins* **11**, 205–217 (1991).
78. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy* page 446 (Springer, New York, 2006).
79. Zondervan, R., Kulzer, F., Orlinskii, S. B. & Orrit, M. Photoblinking of rhodamine 6G in poly(vinyl alcohol): radical dark state formed through the triplet. *J. Phys. Chem. A* **107**, 6770–6776 (2003).
80. Vosch, T. et al. Probing Förster type energy pathways in a first generation rigid dendrimer bearing two perylene imide chromophores. *J. Phys. Chem. A* **107**, 6920–6931 (2003).
81. Hübner, C. G. et al. Photon antibunching and collective effects in the fluorescence of single bichromophoric molecules. *Phys. Rev. Lett.* **91**, 093903 (2003).
82. Antonik, M., Felekyan, S., Gaiduk, A. & Seidel, C. A. M. Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *J. Phys. Chem. B* **110**, 6970–6978 (2006).
83. Nir, E. et al. Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments. *J. Phys. Chem. B* **110**, 22103–22124 (2006).
84. Kaim, G. et al. Coupled rotation within single F₀F₁ enzyme complexes during ATP synthesis or hydrolysis. *FEBS Lett.* **525**, 156–163 (2002).
85. Blatz, A. L. & Magleby, K. L. Correcting single channel data for missed events. *Biophys. J.* **49**, 967–980 (1986).

A stability limit for the atmospheres of giant extrasolar planets

Tommi T. Koskinen¹, Alan D. Aylward¹ & Steve Miller¹

Recent observations of the planet HD209458b indicate that it is surrounded by an expanded atmosphere of atomic hydrogen that is escaping hydrodynamically^{1–3}. Theoretically, it has been shown that such escape is possible at least inside an orbit of 0.1 AU (refs 4 and 5), and also that H_3^+ ions play a crucial role in cooling the upper atmosphere^{5,6}. Jupiter's atmosphere is stable⁷, so somewhere between 5 and 0.1 AU there must be a crossover between stability and instability. Here we show that there is a sharp breakdown in atmospheric stability between 0.14 and 0.16 AU for a Jupiter-like planet orbiting a solar-type star. These results are in contrast to earlier modelling^{4,8} that implied much higher thermospheric temperatures and more significant evaporation farther from the star. (We use a three-dimensional, time-dependent coupled thermosphere–ionosphere model⁶ and properly include cooling by H_3^+ ions, allowing us to model globally the redistribution of heat and changes in molecular composition.) Between 0.2 and 0.16 AU cooling by H_3^+ ions balances heating by the star, but inside 0.16 AU molecular hydrogen dissociates thermally, suppressing the formation of H_3^+ and effectively shutting down that mode of cooling.

We have not attempted to model any specific exoplanet, but rather to create a consistent sequence of models at different orbital distances to characterize the general properties of extrasolar giant planet thermospheres. We model the upper atmosphere as a viscous fluid composed of hydrogen and helium by solving the Navier–Stokes equations of continuity, momentum and energy in spherical pressure coordinates, with a pressure range of 2 μbar to 3.7 pbar. The upper atmospheres of extrasolar giant planets are heated by the absorption of stellar X-ray and ultraviolet (XUV) radiation, which also ionizes the neutral species. We calculate ion densities by assuming photochemical equilibrium and reactions similar to those that take place in Jupiter's thermosphere (Supplementary Table 1). In general, our input parameters correspond to a Jupiter-type gas giant. All our models assume rotational synchronization with the host star, because this is expected for close-in exoplanets. Further details of the model are given in Supplementary Methods.

Modelling and observations indicate that H_3^+ ions act as a critical coolant in gas giant thermospheres in the Solar System⁹. Our simulations between 0.16 and 0.2 AU indicate that significant quantities of H_3^+ ions also form throughout the upper atmosphere of giant exoplanets (Fig. 1). H_3^+ radiates effectively in the infrared and acts as an important coolant in the lower thermosphere but collisions between H_2 and the H_3^+ ions cannot maintain local thermodynamic equilibrium (LTE) at higher altitudes. As an improvement on previous modelling⁶, we therefore performed detailed balance calculations for the dominant vibrational lines of H_3^+ and used the results to adjust the local volume emission rate¹⁰ for non-LTE conditions. We find that the non-LTE cooling rate is almost identical to the LTE rate in the lower thermosphere but that it drops to less than 0.1% of the LTE rate near the top boundary.

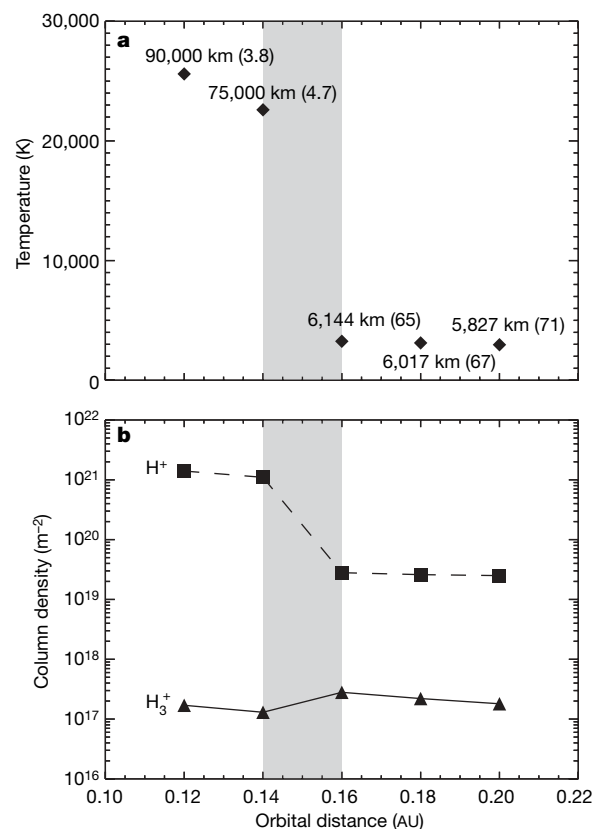


Figure 1 | Temperature and column densities of the dominant ion species versus orbital distance. **a**, The filled diamonds show the globally averaged temperature near the upper boundary at 3.7 pbar. The extent of the thermosphere above the lower boundary altitude is shown next to the data points. The figure in parentheses is the thermal escape parameter¹¹ for atomic hydrogen. The grey-shaded area is the crossover region between stability and hydrodynamic escape. Within this limit all H_2 has been converted into atomic hydrogen in the upper thermosphere (at pressures less than ~ 1 nbar). The data points at 0.12 and 0.14 AU are shown roughly at the onset of hydrodynamic escape and thus do not refer to steady-state models. For the stable models the 3.7 pbar level is near the exobase whereas for the 0.14 and 0.12 AU unstable models the exobase has shifted to a slightly higher altitude. **b**, Substellar column densities of H_3^+ and H^+ (in m^{-2}). Once H_2 is dissociated, H_3^+ no longer forms above the 1 nbar level. Any remaining H_3^+ quickly recombines and the ion is effectively removed from the upper thermosphere. At the same time atomic hydrogen is quickly ionized and the column density increases steeply. Despite its removal from the upper thermosphere, much of the H_3^+ initially survives in the lower thermosphere.

¹Atmospheric Physics Laboratory, Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK.

The circulation depends mainly on the pressure gradient forces induced by non-uniform heating and Coriolis forces arising from the rotation of the planet (Fig. 2). The thermal balance of the model depends overwhelmingly on the absorption of stellar radiation and infrared cooling that regulate the energies available on the dayside of the planet for redistribution by circulation and heating (Fig. 3). We find that between 0.16 and 0.2 AU, H_3^+ infrared cooling almost exactly balances XUV heating, with downward conduction making up for the small difference. In this orbital range the globally averaged temperature around the upper boundary is $\sim 3,000$ K and the extent of the thermosphere above the lower boundary is $\sim 6,000$ km (Figs 1 and 2)—that is, less than 10% of the planetary radius. According to the formal definition of the exobase as the altitude at which the escaping particle undergoes one collision within one pressure scale height, the 3.7 pbar level is near the exobase. At this level the thermal escape parameter¹¹—which determines the stability of the atmosphere—ranges from 65 to 71 between 0.16 and 0.2 AU, implying atmospheric stability and negligible evaporation. (If the escape parameter is below about 30, Jeans escape becomes significant; hydrodynamic escape occurs once the escape parameter is about 1.5 or less.)

As the model is moved inwards from 0.16 AU, however, the character of the upper atmosphere suddenly changes within a narrow range of orbital distances. As the temperature increases, H_2 dissociates, owing

to collisions with other molecules. Once thermal dissociation becomes significant, it happens very rapidly and transport effects bringing more H_2 from below on the dayside are not sufficient to compensate for it. This impedes the formation of H_3^+ and the rapid loss of infrared cooling causes further catastrophic breakdown of the H_2 atmosphere. For example, at 0.14 AU the infrared cooling rate is only $\sim 75\%$ of the XUV heating rate, and downward conduction is not efficient enough to make up for the large difference. Thus, at the onset of hydrodynamic escape at 0.14 AU, roughly $\sim 25\%$ of the absorbed stellar XUV energy is available to power escape. The thermosphere heats up and expands dramatically, producing an inflated upper atmosphere with an extent comparable to or larger than the radius of the planet and temperatures in excess of 20,000 K. At pressures lower than 0.7 nbar, the whole atmosphere turns into atomic hydrogen, most of which is quickly ionized. The thermal kinetic energy becomes comparable to the gravitational potential energy near the upper boundary, implying that the atmosphere begins to escape hydrodynamically¹¹. This condition is reached in our model at the ‘stability limit’ between 0.14 and 0.16 AU, implying that the thermosphere escapes at the upper boundary in bulk and generates a ‘planetary wind’.

We have not attempted to model the inflated extrasolar giant planets after hydrodynamic escape has set in, because some of our basic assumptions become suspect in this regime. For example, the

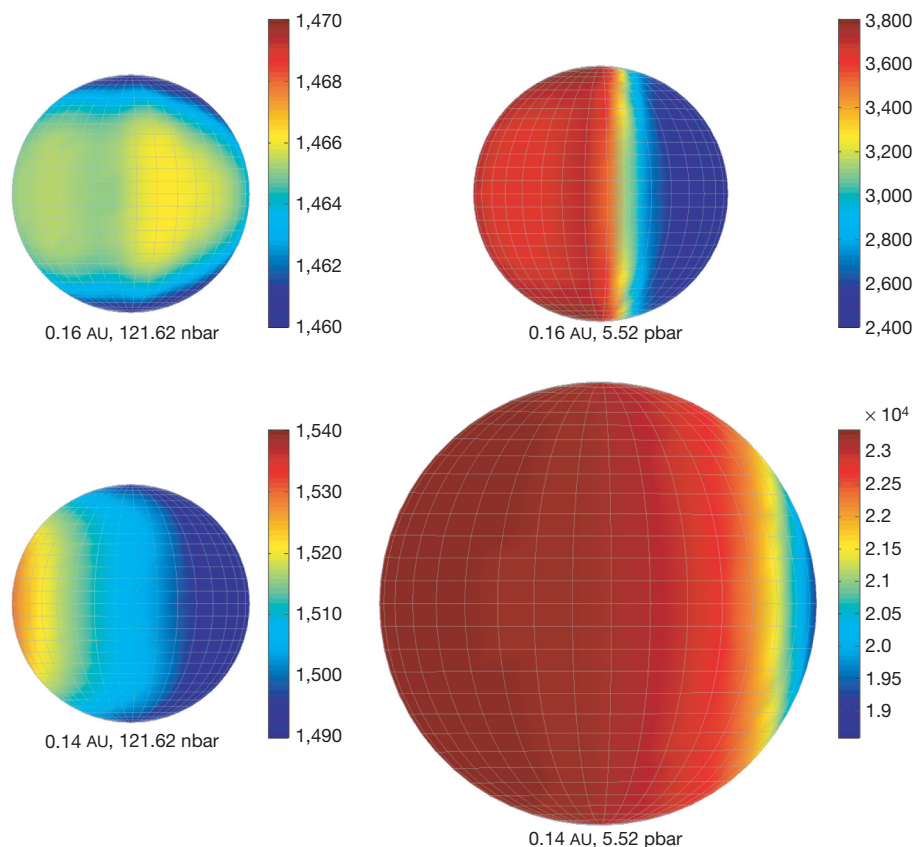


Figure 2 | Hemispheric projections of the horizontal temperature distribution centred at the ‘dusk’ terminator at an orbital distance of 0.16 AU (top) and 0.14 AU (bottom). Two different pressure levels, 122 nbar (left)—corresponding to the bottom of the thermosphere—and 5.5 pbar (right), at the top of the model, are shown. The size of the globes is scaled to the relative planetary radius at the pressure level shown. At 0.16 AU around the 122 nbar level the temperature is nearly uniform between 1,460 and 1,470 K. The altitude is roughly 650 km above the lower boundary of the model. The equatorial wind flows eastward around the planet with a maximum speed of $\sim 120 \text{ m s}^{-1}$, driven by the pressure gradients and the Coriolis force generated by the planet’s orbital motion (with a period of ~ 20 Earth days). The hottest region is located downstream from the substellar point near the dusk terminator. At 5.5 pbar the substellar

temperature is 3,750 K and the antistellar temperature is 2,350 K. Strong winds of over 2 km s^{-1} flow from the dayside to the nightside. Along the equator opposing flows converge near the ‘dawn’ terminator. At 0.14 AU, near the 122 nbar level, the hottest region is at the substellar point, with a temperature of 1,530 K, while the antistellar temperature is 1,480 K. At this level the circulation pattern and altitude are roughly comparable to the 0.16 AU model. The upper thermosphere, however, expands to 75,000 km above the lower boundary, with a substellar temperature of over 23,000 K. In the extended envelope, radiation can penetrate much further into the nightside and so the temperature is roughly uniform horizontally, apart from the small region around the antistellar point where it drops to $\sim 17,000$ K. The resulting day–night winds converge at the antistellar point and the maximum wind speed exceeds 4 km s^{-1} .

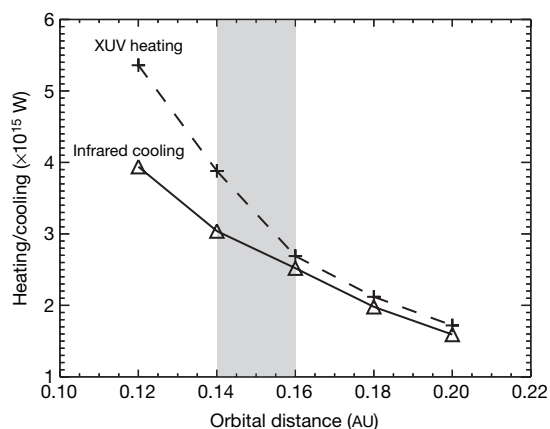


Figure 3 | The total XUV heating and infrared cooling rates at different orbital distances integrated over all pressure levels and both hemispheres. Outside the ‘stability limit’ (grey-shaded area), the infrared cooling from H_3^+ ions effectively balances the XUV heating. The data points at 0.12 and 0.14 AU are shown approximately at the onset of hydrodynamic escape and thus do not refer to steady-state models. However, the XUV heating rate increases more rapidly with decreasing orbital distance than infrared cooling within the limit. At 0.14 AU about 25% of the absorbed XUV energy is available to power escape. The infrared cooling rate keeps increasing because H_3^+ survives initially in the lower thermosphere and the emission rate is sensitive to increasing temperatures.

model assumes hydrostatic equilibrium and that the concentrations of neutral species are not directly affected by ionization. But in the inflated thermosphere H^+ is the dominant species, rendering the latter assumption questionable. The former assumption is made dubious by the fact that hydrodynamic outflow implies vertical accelerations. We note that bulk outflow will result in adiabatic cooling and cooling by vertical advection that may change the temperatures and composition in a self-consistent, hydrodynamically escaping model atmosphere. There is, however, qualitative agreement with the implications of the observations of HD209458b and our results, because we predict that the planet has an expanded thermosphere that is undergoing hydrodynamic escape. A recent one-dimensional, non-hydrostatic model of HD209458b indicates that H_3^+ does not form in the upper thermosphere due to dissociation of H_2 at high temperatures⁵. Instead, the dominant species at high altitudes is H^+ and the outer layers of the atmosphere extend to several planetary radii.

The ‘ H_2 -breakdown limit’ to thermospheric stability depends on the stellar XUV flux. The high-energy XUV emissions of the Sun originate in the solar chromosphere, transition region and the corona and are due to the release of magnetic dynamo-generated energy¹². The strength of the dynamo is determined by the rotation rate of the star. Current observational evidence indicates that young solar-type stars rotate much faster than the current Sun and that the rotation rate decreases during their evolution in the main sequence. The

implication is that the XUV emissions from young solar-type stars are higher than those used in this study.

To estimate the response of our model to XUV emissions from solar-type stars of different ages we used six stars from the ‘Sun in Time’ program, which uses solar-type stars of different ages to characterize the evolution of the Sun’s XUV emissions^{4,12}. We then used a simple scaling law to estimate the orbital distance around these stars where the total XUV flux would be similar to the total XUV flux received near our ‘stability limit’ at 0.16 AU around the Sun (Table 1). Between 100 Myr and 6.7 Gyr the limit moves from ~ 1.7 to 0.13 AU as the flux decreases by a factor of ~ 180 . Similar scaling can be used to estimate the location of the stability limit around stars of different spectral types once the XUV fluxes are known. Our results disagree with earlier work predicting higher temperatures within 1 AU and blow-off within 0.3 AU for current solar XUV fluxes⁴. The escape rates derived from these estimates are likely to be too high because they do not account for infrared cooling or circulation. Even in the expanded atmosphere considerable amounts of H_3^+ survive in the lower thermosphere.

Some estimates of the extrasolar giant planet escape rates are based on the idea that all absorbed EUV energy goes into powering escape⁸. We note that this approach is only appropriate for planets orbiting within our stability limit. Further out from the limit almost none of the absorbed energy is available for escape. Instead heating is balanced by infrared cooling and redistributed around the planet by circulation.

Combined evolutionary and atmospheric models show that the survival of a giant exoplanet depends on its initial mass and the evaporation rates. If the initial mass is lower than a critical value, which depends on the evaporation and contraction timescales, the planet will evaporate entirely in less than 5 Gyr (ref. 13). However, few if any attempts have been made to model the combined effects of migration and atmospheric evolution. Migration rates currently thought reasonable vary from 0.1 AU Myr^{-1} (ref. 14) to 10 AU Myr^{-1} (R. P. Nelson and M. Fogg, personal communication). Even with the slower rate, a giant planet forming between 5 and 20 AU from the Sun (where the majority of giant planets form) and then migrating inwards would reach the relevant stability limit of 1.7 AU during the first 33 to 183 Myr (Table 1). This, however, does not challenge the conclusion suggested by various groups that close-in extrasolar giant planets appear to be stable against evaporation^{5,8,15} during the lifetime of a main sequence star, because our work implies that escape rates are lower than those anticipated by earlier one-dimensional models that neglect infrared cooling.

We have calculated the total emitted output power from H_3^+ ions at different orbital distances (Fig. 3). Our model is also capable of calculating signal strengths for specific vibrational lines of H_3^+ . The predicted signals are weak, and detection with current technology would prove challenging¹⁶. We do encourage, however, an attempt to detect H_3^+ on a giant exoplanet as soon as suitable technology becomes available, because such detection would resolve many of the existing uncertainties regarding extrasolar giant planet atmospheres.

Received 29 June; accepted 4 October 2007.

Table 1 | The Sun in Time targets and thermospheric stability limits

Name	HD	Type	Age (Gyr)	F ($\text{erg s}^{-1} \text{cm}^{-2}$)	Limit (AU)
EK Dra	129333	G1.5 V	0.1	513.5	1.68
p ¹ UMa	72905	G1.5 V	0.3	129.3	0.84
k ¹ Cet	20630	G5 V	0.65	51.1	0.53
b Com	114710	G0 V	1.6	16.0	0.30
The Sun	G2 V	4.6	4.64	0.16
b Hyi	2151	G2 IV	6.7	2.9	0.13

The total XUV flux F applies to the 0.1–118 nm wavelength interval and is normalized to a distance of 1 AU from the star. The observations of the Sun in Time program cover a range of wavelengths from 0.1 to 170 nm, excluding the gap between 36 and 92 nm, which is affected by very strong absorption by the interstellar medium due to the H I Lyman α continuum. In this gap, the irradiances have been estimated by comparison with the flux evolution in other wavelength regions. See ref. 12 for further details of the stars in the sample and the data reduction process.

1. Vidal-Madjar, A. *et al.* An extended upper atmosphere around the extrasolar planet HD209458b. *Nature* **422**, 143–146 (2003).
2. Vidal-Madjar, A. *et al.* Detection of oxygen and carbon in the hydrodynamically escaping atmosphere of the extrasolar planet HD209458b. *Astrophys. J.* **604**, L69–L72 (2004).
3. Ballester, G., King, D. K. & Herbert, F. The signature of hot hydrogen in the atmosphere of the extrasolar planet HD 209458b. *Nature* **445**, 511–514 (2007).
4. Lammer, H. *et al.* Atmospheric loss of exoplanets resulting from stellar X-ray and extreme-ultraviolet heating. *Astrophys. J.* **598**, L121–L124 (2003).
5. Yelle, R. V. Aeronomy of extra-solar giant planets at small orbital distances. *Icarus* **170**, 167–179 (2004).
6. Koskinen, T. T., Aylward, A. D., Smith, C. G. A. & Miller, S. A thermospheric circulation model for extrasolar giant planets. *Astrophys. J.* **661**, 515–526 (2007).

7. Strobel, D. F. in *Atmospheres in the Solar System, Comparative Aeronomy* (eds Mendillo, M., Nagy, A. & Waite, J. H.) 7–22 (AGU, Washington DC, 2002).
 8. Lecavelier des Etangs, A. A diagram to determine the evaporation status of extrasolar planets. *Astron. Astrophys.* **461**, 1185–1193 (2007).
 9. Miller, S. *et al.* H_3^+ : the driver of giant planet atmospheres. *Phil. Trans. R. Soc. A* **364**, 3121–3137 (2006).
 10. Neale, L., Miller, S. & Tennyson, J. Spectroscopic properties of the H_3^+ molecule: a new calculated line list. *Astrophys. J.* **464**, 516–520 (1996).
 11. Hunten, D. M. The escape of light gases from planetary atmospheres. *J. Atmos. Sci.* **30**, 1481–1494 (1973).
 12. Ribas, I., Guinan, E. F., Gudel, M. & Audard, M. Evolution of the solar activity over time and effects on planetary atmospheres. I. High-energy irradiances (1–1700 Å). *Astrophys. J.* **622**, 680–694 (2005).
 13. Baraffe, I. *et al.* The effect of evaporation on the evolution of close-in giant planets. *Astron. Astrophys.* **419**, L13–L16 (2004).
 14. Rice, W. K. M. & Armitage, P. J. Quantifying orbital migration from exoplanet statistics and host metallicities. *Astrophys. J.* **630**, 1107–1113 (2005).
 15. Tian, F., Toon, O. B., Pavlov, A. A. & De Sterck, H. Transonic hydrodynamic escape of hydrogen from extrasolar planetary atmospheres. *Astrophys. J.* **621**, 1049–1060 (2005).
 16. Shkolnik, E. E., Gaidos, E. & Moskovitz, N. No detectable H_3^+ emission from the atmospheres of hot Jupiters. *Astron. J.* **132**, 1267–1274 (2006).
- Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.
- Acknowledgements** T.T.K. has been supported by UCL and Perren studentships and A.D.A and S.M. have been supported by the Science and Technology Facilities Council (STFC). This work was partly carried out on the Keter High Performance Computer System, which is managed by the Miracle Astrophysics Project and funded by STFC. SOLAR2000 Research Grade v2.23 irradiances are provided by Space Environment Technologies. We also thank R. Nelson and M. Fogg for advice on giant planet migration rates and N. Achilleos for assistance with global plots.
- Author Contributions** T.T.K. developed the thermospheric circulation model for extrasolar giant planets based on existing models of gas giants in the Solar System. T.T.K, A.D.A and S.M. directed and carried out the modelling. S.M. calculated the non-LTE emission rates for H_3^+ .
- Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.T.K. (tommi@apl.ucl.ac.uk).

Cryptomare magmatism 4.35 Gyr ago recorded in lunar meteorite Kalahari 009

Kentaro Terada¹, Mahesh Anand^{2,3}, Anna K. Sokol⁴, Addi Bischoff⁴ & Yuji Sano⁵

The origin and evolution of the Moon remain controversial^{1,2}, with one of the most important questions for lunar evolution being the timing and duration of basaltic (mare) magmatism^{1,3–8}. Here we report the result of ion microprobe U–Pb dating of phosphates in a lunar meteorite, Kalahari 009, which is classified as a very-low-Ti mare-basalt breccia. *In situ* analyses of five phosphate grains, associated with basaltic clasts, give an age of 4.35 ± 0.15 billion years. These ancient phosphate ages are thought to represent the crystallization ages of parental basalt magma, making Kalahari 009 one of the oldest known mare basalts. We suggest that mare basalt volcanism on the Moon started as early as 4.35 Gyr ago, relatively soon after its formation and differentiation, and preceding the bulk of lunar volcanism which ensued after the late heavy bombardment around 3.8–3.9 Gyr (refs 7 and 8). Considering the extremely low abundances of incompatible elements such as thorium and the rare earth elements in Kalahari 009 (ref. 9) and recent remote-sensing observations illustrating that the cryptomaria tend to be of very-low-Ti basalt type^{10–12}, we conclude that Kalahari 009 is our first sample of a very-low-Ti cryptomare from the Moon.

Chronological studies of numerous returned samples of mare basalt and related pyroclastic deposits from the Moon have been well documented (for summaries, see refs 6 and 7). In general, the high-Ti basalts from the Apollo 11 and Apollo 17 sites are relatively old; commonly ranging in age from 3.5 to 3.9 billion years (Gyr). In contrast, low-Ti mare basalt samples are generally younger; they range in age from 3.1 to 3.4 Gyr, although some exceptions exist (Apollo 14 mare basalts have ages from 3.9 to 4.2 Gyr; refs 13 and 14). On the basis of the available data for very-low-Ti (VLT) basalts, Nyquist *et al.*⁶ have suggested that the formation ages of VLT mare basalts from the site Luna 24 are younger, at 3.2–3.3 Gyr. In addition, recent chronological studies of lunar meteorites provide additional insights into the thermal histories of the lunar interior. Although these meteorites are possibly derived from unexplored regions of the Moon, they are mainly the products of later magmatic activity, spanning a billion years, from 3.9 to 2.9 Gyr (refs 15–17). Thus, the majority of known mare basalt samples have crystallization ages younger than about 3.9 Gyr, leading to the hypothesis that mare volcanism occurred mainly after the late heavy bombardment around 3.8–3.9 Gyr.

Two previous studies have reported ~ 4.2 Gyr ages for basalt clasts in Apollo 14 breccias^{13,14} (exclusion of some data points in one case yielded an older age, possibly up to 4.33 Gyr; ref. 14), suggesting that mare volcanism occurred early during the lunar crust formation. Similarly, VLT basalts found as fragments in lunar sample Breccia 72235 must be older than the melt-rich matrix of the breccia, which crystallized at 4.01 Gyr (ref. 18). Thus, although we may have samples

of some ancient mare basalts, we still lack definitive evidence for lunar basalts formed before 4.0–4.2 Gyr, and more importantly, those samples that can be considered products of a period in the lunar history during which cryptomare deposits (mare deposits hidden from direct view by superposed impact craters and bright ejecta¹⁹) were emplaced.

Kalahari 009 is a fragmental basaltic lunar meteorite collected from Botswana in 1999 as a single rock of about 13.5 kg (ref. 20). In terms of mineralogy and bulk chemical composition, Kalahari 009 mainly consists of VLT mare basalt components. The lunar heritage of this rock is confirmed by the bulk-oxygen isotope composition, mineralogy, bulk-rock elemental composition, and Fe/Mn ratios in olivine and pyroxenes²⁰. The rock consists of fragments of basaltic clasts set in a fine-grained matrix. Many of the basaltic clasts have a coarse-grained subophitic texture. Clasts and matrix display the same composition. Pyroxene is the most abundant phase, followed by plagioclase. Pyroxene grains are zoned with compositions from pigeonite to augite ($\text{En}_{3-64}\text{Fs}_{28-59}\text{Wo}_{7-41}$) and commonly display exsolution lamellae at the scale of $<5 \mu\text{m}$ (ref. 20). Most feldspars are anorthitic ($\text{An} > 90$), but some are more sodic. Olivine is Fe-rich (Fa_{50-100})²⁰ and symplectite intergrowths of hedenbergite + fayalite + SiO_2 were frequently observed. Similar symplectite assemblages have been reported from other lunar basalts¹⁷. Minor mineral phases in the rock include ilmenite, spinel, troilite, FeNi metal, phosphates and Zr-bearing phase (possibly zirconolite), all typical of mare basalts. Most phosphate grains are of the order of $10 \times 10 \mu\text{m}$ in dimension and are always associated with late-stage fractionates (for example, ilmenites, Fe-rich olivines and pyroxenes), similar to those seen in other mare basalts (Fig. 1, Table 1).

In addition to these characteristics, Kalahari 009 possesses some unique geochemical features. (1) Although, in terms of bulk-rock major-element composition, it is similar to some Apollo 14 aluminous mare basalts, it has the lowest bulk-rock rare-earth-element and Th abundances of all mare basalt samples: 0.1 p.p.m. Th (ref. 9). (2) It is almost entirely composed of mare-basaltic material; and until now no highlands-derived material has been identified in Kalahari 009, although it has been argued that an anorthositic breccia, Kalahari 008, was formed in the vicinity of Kalahari 009 and that both rocks were ejected from the Moon as one polymict meteoroid²⁰. (3) The Ar–Ar age of 1.70 ± 0.04 Gyr has been suggested to be the crystallization age of Kalahari 009 (ref. 21), but this extremely young age is probably a result of Ar-loss during a post-crystallization impact process⁹. Evidence of shock is present in Kalahari 009 in terms of maskelynite, formed by solid-state transformation of plagioclase.

For a better understanding of the thermal history recorded in Kalahari 009, we carried out *in situ* U–Pb dating of phosphate minerals using an ion microprobe. In lunar basalts, phosphates are the main carriers of U and are also resistant to secondary thermal events

¹Department of Earth and Planetary Systems Science, Hiroshima University, Higashi-Hiroshima 739-8526, Japan. ²Department of Earth and Environmental Sciences, CEPSAR, Walton Hall, The Open University, Milton Keynes, MK7 6AA, UK. ³Department of Mineralogy, The Natural History Museum, Cromwell Road, London, SW7 5BD, UK. ⁴Institut für Planetologie, Wilhelm-Klemm-Strasse 10, 48149 Münster, Germany. ⁵Center for Advanced Marine Research, Ocean Research Institute, The University of Tokyo, Nakano-ku, Tokyo 164-8639, Japan.

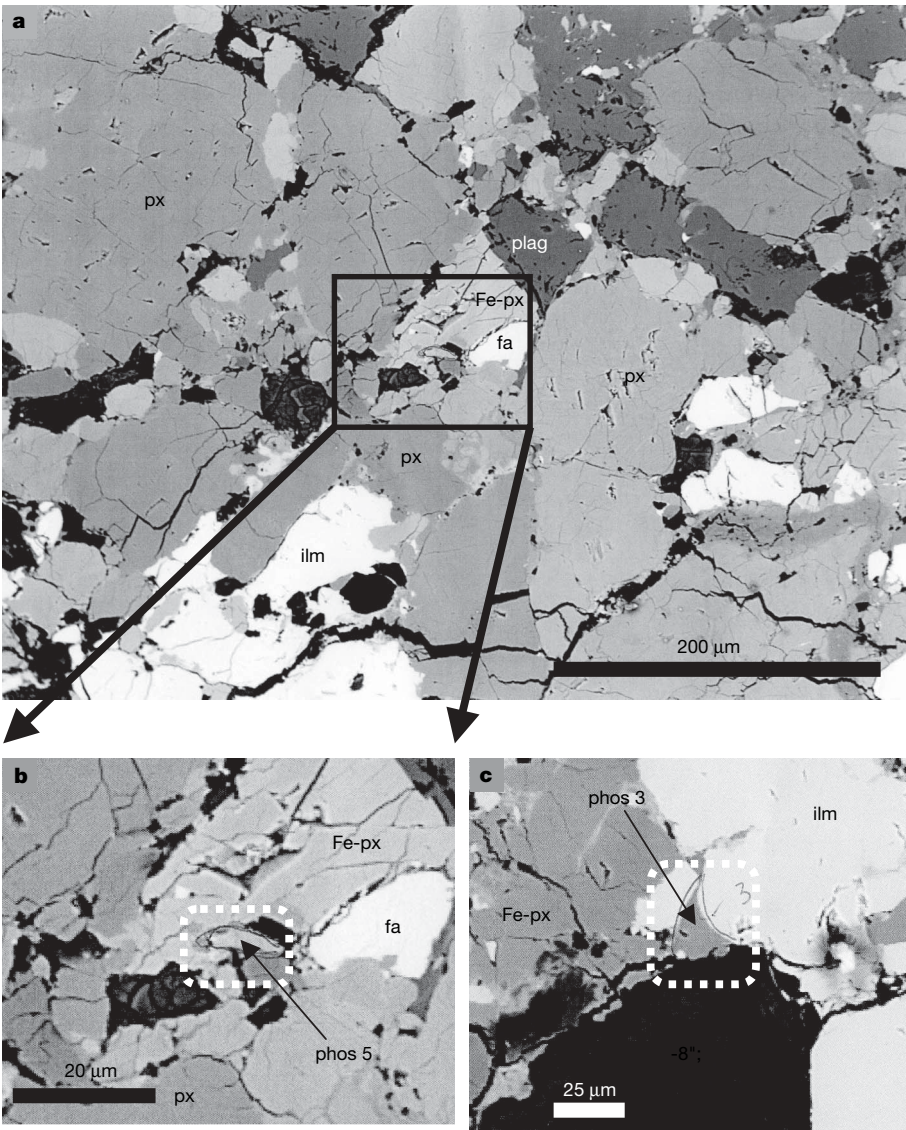


Figure 1 | Back-scattered electron images of areas containing phosphate grains in Kalahari 009. **a**, These phosphate grains are typically associated with minerals belonging to the assemblage: Fe-pyroxene (Fe-px), ilmenite (ilm) and fayalite (fa). Other main minerals in the rock such as plagioclase (plag) and more magnesian pyroxenes (px) are also labelled. **b**, A close-up of the phosphate grain 5 (phos 5), shown in **a**. **c**, Close-up of phosphate grain 3 (phos 3).

owing to their relatively high closure temperatures of ~500–600 °C for Pb. We have previously demonstrated the robustness of our U–Pb *in situ* dating technique for dating crystallization and alteration ages of phosphates in lunar basaltic rocks^{15–17}. For phosphates in Kalahari 009, we obtained an U–Pb isochron age of $4,251 \pm 750$ million years (Myr) and a Pb–Pb isochron age of $4,368 \pm 160$ Myr (Figs 2a, b and Table 2), suggesting that the U–Pb systematics agree within the analytical uncertainties at the 95% confidence level. The total Pb/U isochron age in the $^{238}\text{U}/^{206}\text{Pb}$ – $^{207}\text{Pb}/^{206}\text{Pb}$ – $^{204}\text{Pb}/^{206}\text{Pb}$ three-dimensional space results in $4,349 \pm 150$ Myr (Fig. 2c). All these ages suggest that the formation age of phosphates in Kalahari

009 is ~4.3 to 4.4 Gyr; when combined with textural, mineralogical, and petrological evidence that these phosphate grains are genetically related to basaltic components in Kalahari 009, the crystallization ages of phosphates directly provide the crystallization age of the basaltic magma, parental to Kalahari 009. Recently, a preliminary Lu–Hf isochron study has also reported a similar ancient crystallization age for Kalahari 009 (ref. 9). Therefore, we propose that Kalahari 009 is one of the oldest-known mare basaltic samples in our collection. Furthermore, the ancient age of Kalahari 009 suggests that the basaltic volcanism on the Moon started relatively soon after its formation and differentiation.

Table 1 | Major-element compositions of phosphates in Kalahari 009

Grain number	Mineral	P ₂ O ₅ (wt%)	CaO (wt%)	MgO (wt%)	FeO (wt%)	Al ₂ O ₃ (wt%)	MnO (wt%)	SiO ₂ (wt%)	TiO ₂ (wt%)	Y ₂ O ₃ (wt%)	SrO (wt%)	F (wt%)	Cl (wt%)	–O = F, Cl	Total
1	Apatite	39.76	51.30	0.34	2.14	0.44	0.04	3.56	0.03	0.60	0.38	2.64	0.67	1.26	100.6
2	Apatite	40.77	51.97	0.06	1.27	0.14	0.02	2.50	0.04	0.77	0.13	3.70	0.53	1.68	100.2
3	Merrillite/apatite	40.73	49.23	0.02	5.29	0.14	0.05	1.44	0.66	1.13	0.12	3.01	0.17	1.31	100.7
4	Merrillite	37.12	45.42	0.24	6.16	0.98	0.04	5.87	0.08	1.45	0.26	2.87	0.28	1.27	99.5
5	Merrillite	37.38	45.96	0.58	6.68	0.68	0.05	6.10	0.19	0.49	0.09	2.14	0.20	0.95	99.6

–O is oxygen equivalent for the amount of chlorine and fluorine present in the mineral.

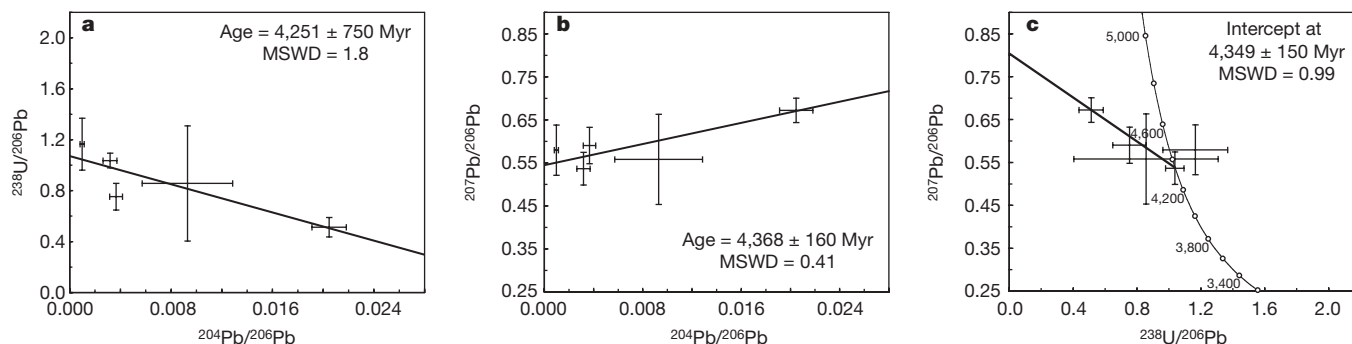


Figure 2 | The result of *in situ* U–Pb dating of phosphates in Kalahari 009. **a**, Inverse U–Pb isochron diagram. **b**, Inverse Pb–Pb isochron diagram. **c**, The projected diagram onto the $^{238}\text{U}/^{206}\text{Pb}$ – $^{207}\text{Pb}/^{206}\text{Pb}$ plane of the total/Pb isochron in the $^{238}\text{U}/^{206}\text{Pb}$ – $^{207}\text{Pb}/^{206}\text{Pb}$ – $^{204}\text{Pb}/^{206}\text{Pb}$ three-dimensional space.

Recent remote-sensing observations have provided some constraints on the relative chronology of the unexplored regions of the Moon, indicating that mare basalt magmatism peaked between 4.0 and 3.8 Gyr (refs 7 and 8) and extended to approximately 1.2–0.9 Gyr (refs 1, 7 and 8). Moreover, the global chemical mapping data suggests that the majority of the observed mare basalt deposits on the lunar surface are enriched in Th (with the highest enrichments seen in nearside deposits), usually considered to be a signature associated with the Procellarum KREEP (enriched in potassium, rare earth element, and phosphorus) terrain magmatism²². These features are quite different from those of Kalahari 009. Schultz and Spudis²³ proposed an alternative hypothesis for the existence of hidden contiguous mafic bodies from which such ancient basalt clasts may have originated, which was later fully confirmed with Earth-based telescopic spectral data²⁴. Subsequently, ref. 19 identified these hidden mafic deposits, which they termed “cryptomaria”. From a global study of all previously identified and suspected hidden mafic deposits, ref. 25 estimated possible age ranges from 3.2 to 4.1 Gyr, suggesting that mare basalt magmatism had started by at least 4.1 Gyr ago on the Moon, as previously proposed on the basis of radiometric age dating of Apollo samples^{13,14}. Recently, on the basis of the spectral and chemical data obtained for dark-haloed craters that are believed to have excavated mare basalts, refs 10–12 have suggested that mare basalts in cryptomaria tend to be of the VLT type, consistent with Kalahari 009 data. Therefore, on the basis of ancient crystallization age and the bulk-rock geochemical characteristics of Kalahari 009, we conclude that it represents our first sample from such VLT-type cryptomaria on the Moon.

The evolution of the lunar highland crust, as determined by returned samples from the Moon, is commonly considered to have involved two distinct stages²⁶. The first stage was the formation of anorthositic to leuconoritic crust as flotation cumulates from an incipient lunar magma ocean at ~4.4–4.5 Gyr, followed by a second stage involving the modification of this early crust from 4.4 to 3.9 Gyr through the crystallization of mafic and ultramafic plutonic rocks such as the highland-magnesian-suite and highland-alkali-suite rocks. Therefore, the important point to note is that the VLT mare-type volcanism occurred at least as early as 4.35 Gyr ago, and just after the first stage of lunar crust formation, but in time it

The uncertainties of plotted data and obtained ages are reported at the 1 σ level and the 95% confidence level, respectively. All data suggest that the crystallization ages of phosphates are about 4.35 Gyr. The linear regressions were calibrated using Isoplot/Ex (see Methods). MSWD, mean square of weighted deviated.

overlapped with the second stage and KREEP magmatism. Unlike highland-magnesian-suite parental magma²⁷ and trace-element-enriched Apollo 14 mare basalts²⁸, the parent magma of Kalahari 009 was derived from a relatively depleted source, and during its ascent from the lunar interior it did not assimilate any KREEP, suggesting that KREEP might not be a global layer around the crust–mantle boundary of the Moon, or at least that it had not formed before Kalahari 009 magmatism.

So far, numerous models have been proposed for the origin and evolution of lunar basalts (for a recent review, see ref. 29), strongly relying on the data available for the timing of onset of mare basalt volcanism and the heat source for mantle melting. One endmember model invokes a ‘passive’ mechanism for lunar magmatism wherein internal heating by radioactive elements associated with KREEP causes partial melting of deeper parts of the Moon³. Another model involving a ‘passive’ mechanism for lunar magmatism is based on large-scale overturn of an initially unstable stratified lunar interior, causing mare magmatism several hundred million years later, after lunar magma ocean crystallization⁴. An alternative model involving an ‘active’ mechanism proposes a two-stage model for lunar magma generation triggered by impacts⁵. In the first stage, a basin-forming impact event causes large degrees of instantaneous mantle melting, followed by a second stage involving adiabatic decompression melting and extrusion of mare basalts into the newly created impact basin⁵. Our findings of ancient mare magmatism (cryptomare) associated with relatively low incompatible element (that is, heat-producing elements) abundances seem compatible with the ‘active’ mechanism model described above. Thus, Kalahari 009 basaltic magmatism could be a manifestation of an ancient basin-forming event at 4.35 Gyr on the Moon.

METHODS SUMMARY

For *in situ* U–Pb dating we used a sensitive high-resolution ion microprobe (SHRIMP) installed at the Hiroshima University, Japan. A 0.3 nA O_2^- primary beam with acceleration voltage of 10 kV was focused to sputter an area ~5 μm in diameter on the phosphate grains and the positive secondary ions were extracted. The mass resolution was set to 5,800 at ^{208}Pb for U–Pb analyses. The abundance ratios of ^{238}U to ^{206}Pb for phosphates were obtained from the observed $^{238}\text{U}^{+}/^{206}\text{Pb}^{+}$ ratios, using an empirical quadratic relationship between the $^{206}\text{Pb}^{+}/^{238}\text{U}^{+}$ and $^{238}\text{U}^{16}\text{O}^{+}/^{238}\text{U}^{+}$ ratios of standard apatite³⁰ derived from

Table 2 | U concentrations and isotope ratios in phosphates for Kalahari 009

Spot number	Mineral	U (p.p.m.)	$^{238}\text{U}/^{206}\text{Pb}$	$^{207}\text{Pb}/^{206}\text{Pb}$	$^{204}\text{Pb}/^{206}\text{Pb}$
KAL01.1	Apatite	195	1.1650 ± 0.2032	0.5796 ± 0.0583	0.000981 ± 0.000168
KAL02.1	Apatite	46	1.0365 ± 0.0575	0.5365 ± 0.0379	0.003179 ± 0.000545
KAL03.1	Merrillite/apatite	123	0.8568 ± 0.4518	0.5582 ± 0.1050	0.009288 ± 0.003568
KAL04.1	Merrillite	76	0.7530 ± 0.1049	0.5903 ± 0.0422	0.003665 ± 0.000508
KAL05.1	Merrillite	126	0.5130 ± 0.0762	0.6722 ± 0.0285	0.020467 ± 0.001350

Uncertainties assigned to the isotopic and elemental ratio are 1 σ , estimated by counting statistics and calibration. Uncertainties of concentration are $\pm 30\%$, estimated by repeated measurements of standard apatite, PRAP.

an alkaline rock of the Prairie Lake circular complex in the Canadian Shield: $1,156 \pm 45$ Myr ($\pm 2\sigma$).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 June; accepted 2 October 2007.

- Schultz, P. H. & Spudis, P. D. Beginning and end of lunar mare volcanism. *Nature* **302**, 233–236 (1983).
- Yin, Q. et al. A short timescale for terrestrial planet formation from Hf–W chronometry of meteorites. *Nature* **418**, 949–951 (2002).
- Wieczorek, M. A. & Phillips, R. J. The “Procellarum KREEP Terrane”: Implications for mare volcanism and lunar evolution. *J. Geophys. Res.* **105**, 20417–20430 (2000).
- Hess, P. C. & Parmentier, E. M. A model for the thermal and chemical evolution of the Moon's interior: Implications for the onset of mare volcanism. *Earth Planet. Sci. Lett.* **134**, 501–514 (1995).
- Elkins-Tanton, L. T., Hager, B. H. & Grove, T. L. Magmatic effects of the lunar late heavy bombardment. *Earth Planet. Sci. Lett.* **222**, 17–27 (2004).
- Nyquist, L. E., Bogard, D. D. & Shih, C.-Y. *The Century of Space Science 1325–1376* (eds Bleeker, J. A. M., Geiss, J. & Huber, M. C. E.) (Kluwer Academic, Dordrecht, 2001).
- Stöffler, D. & Ryder, G. Stratigraphy and isotope ages of lunar geologic units: chronological standard for the inner solar system. *Space Sci. Rev.* **96**, 9–54 (2001).
- Hiesinger, H., Head, J. W., Wolf, U., Jaumann, R. & Neukum, G. Ages and stratigraphy of mare basalts in Oceanus Procellarum, Mare Nubium, Mare Cognitum, and Mare Insularum. *J. Geophys. Res.* **108** (1) 1–27 (2003).
- Schulz, T. et al. Chemical composition and Lu/Hf-age of the Lunar mare basalt meteorite Kalahari 009. *Meteorit. Planet. Sci.* **42**, A137 (2007).
- Giguere, T. A. et al. Remote sensing studies of the Lomonosov-Fleming region of the Moon. *J. Geophys. Res.* **108** (4) 1–14 (2003).
- Hawke, B. R. et al. Remote sensing and geologic studies of the Balmer-Kapteyn region of the Moon. *J. Geophys. Res.* **110**, E06004, doi:10.1029/2004JE002383 (2005).
- Hawke, B. R. et al. The earliest mare basalts. *Lunar Planet. Sci. Conf.* **36**, abstr. 1642 (2005).
- Taylor, L. A. et al. Pre-4.2 AE mare-basalt volcanism in the lunar highlands. *Earth Planet. Sci. Lett.* **66**, 33–47 (1983).
- Dasch, E. J., Shih, C.-Y., Bansal, B. M., Wiesmann, H. & Nyquist, L. E. Isotopic analysis of basaltic fragments from lunar breccia 14321—chronology and petrogenesis of pre-Imbrium mare volcanism. *Geochim. Cosmochim. Acta* **51**, 3241–3254 (1987).
- Terada, K. et al. Uranium-lead systematics of phosphates in lunar basaltic regolith breccia, Meteorite Hills 01210. *Earth Planet. Sci. Lett.* **259**, 77–84 (2007).
- Anand, M. et al. Petrology and geochemistry of LaPaz icefield 02205: a new unique low-Ti mare-basalt meteorite. *Geochim. Cosmochim. Acta* **70**, 246–264 (2006).
- Anand, M. et al. Petrogenesis of lunar meteorite EET 96008. *Geochim. Cosmochim. Acta* **67**, 3499–3518 (2003).
- Leich, D. A., Kahl, S. B., Kirschbaum, A. R., Niemeyer, S. & Phinney, D. Rare gas constraints on the history of Boulder 1, Station 2, Apollo 17. *The Moon* **14**, 407–444 (1975).
- Head, J. W. III & Wilson, L. Lunar mare volcanism: Stratigraphy, eruption conditions, and the evolution of secondary crusts. *Geochim. Cosmochim. Acta* **56**, 2155–2175 (1992).
- Sokol, A. K. & Bischoff, A. Meteorites from Botswana. *Meteorit. Planet. Sci. A* **40**, 177–184 (2005).
- Fernandes, V. A., Burgess, R., Bischoff, A., Sokol, A. K. & Haloda, J. Kalahari 009 and north east Africa 003: young (<2.5 Ga) lunar mare basalt. *Lunar Planet. Sci. Conf.* **38**, abstr. 1611 (2007).
- Gillis, J. J., Jolliff, B. L. & Korotev, R. L. Lunar surface geochemistry: Global concentrations of Th, K, and FeO. *Geochim. Cosmochim. Acta* **68**, 3791–3805 (2004).
- Schultz, P. H. & Spudis, P. D. Evidence for ancient mare volcanism. *Proc. Lunar Planet. Sci. Conf.* **10**, 2899–2918 (1979).
- Hawke, B. R. & Bell, J. F. Remote sensing studies of lunar dark-halo impact craters: Preliminary results and implications for early volcanism. *Proc. Lunar Planet. Sci. Conf.* **12**, 665–678 (1981).
- Antonenko, I. Global estimates of cryptomare deposits: implications for lunar volcanism. *Lunar Planet. Sci. Conf.* **30**, abstr. 1703 (1999).
- Snyder, G. A. et al. in *Origin of the Earth and Moon* (eds Canup, R. M. & Righter, K.) 361–395 (Univ. Arizona Press, Tucson, 2000).
- Papike, J. J. et al. Evolution of the lunar crust: SIMS study of plagioclase from ferroan anorthosites. *Geochim. Cosmochim. Acta* **61**, 2343–2350 (1997).
- Dickinson, T. et al. Apollo 14 aluminous mare basalts and their possible relationship to KREEP. *Proc. Lunar Planet. Sci. Conf.* **15** (2) C365–C374 (1985).
- Shearer, C. K. et al. Thermal and magmatic evolution of the moon. *Rev. Mineral. Geochem.* **60**, 365–518 (2006).
- Sano, Y., Oyama, T., Terada, K. & Hidaka, H. Ion microprobe U–Pb dating of apatite. *Chem. Geol.* **153**, 249–258 (1999).

Acknowledgements We thank K. R. Ludwig for providing the Isoplot/Ex program for U–Pb age calibration. We thank A. G. Tindle and J. Berndt-Gerdes for assistance with microprobe work. We thank L. E. Thomas and R. A. Spicer for proof-reading the manuscript for English usage. We also thank L. E. Nyquist and J. Head for their constructive comments. This contribution is an outcome of a joint project between the Hiroshima University and the Open University. This study is partly supported by a Scientific Research Grant of the Ministry of Education, Culture, Sports, Science and Technology, the Itoh Science foundation and the German Research Foundation. M.A. acknowledges support of PPARC and RCUK fellowships.

Author Contributions K.T. and M.A. contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.T. (terada@sci.hiroshima-u.ac.jp) or M.A. (M.Anand@open.ac.uk).

METHODS

After the mineralogical investigations of the Kalahari 009 thin-section, using an electron probe micro-analyser, the thin-section was re-polished slightly using 0.25 μm diamond paste. Then, after cleaning to minimize surface contaminant Pb, the section was gold-coated to prevent charging of the sample surface during SHRIMP analyses. To further reduce an already very small $x-1\text{PbH}^+$ interference on the $^3\text{Pb}^+$ peaks, the thin-section was evacuated in the sample lock overnight. An important final step before the actual analysis involved the rastering of the primary ion beam over the entire sample surface for 3 min to remove any remaining possible contaminants.

A 0.3 nA O_2^- primary beam with acceleration voltage of 10 kV was focused to sputter an area $\sim 5\text{ }\mu\text{m}$ in diameter on the phosphate grains, and positive secondary ions were extracted and detected on a single electron multiplier by peak switching. The mass resolution was set at 5,800 at ^{208}Pb for U–Pb analyses. The magnet was cyclically peak-stepped from mass 159 ($^{40}\text{Ca}^{31}\text{P}^{16}\text{O}_3^+$) to mass 254 ($^{238}\text{U}^{16}\text{O}^+$), including background, all Pb isotopes, and masses 238 for ^{238}U . No significant isobaric interferences were detected in this mass range for the phosphates (for example, the mass peak of ^{159}Tb (158.925 atomic mass units, AMU) is clearly separated from that of $^{40}\text{Ca}^{231}\text{P}^{16}\text{O}_3^+$ (158.884 AMU) at the mass resolution of 5,800). Further experimental details of the U–Pb analysis and the calibration of the data have been described elsewhere³⁰. For age calculations, we used the Isoplot/Ex program³¹.

31. Ludwig, K. R. *Users Manual for Isoplot/Ex: a Geochronological Toolkit for Microsoft Excel* (Berkeley Geochronology Center, Special Publication 1a, 2001).

Low-temperature shear modulus changes in solid ^4He and connection to supersolidity

James Day¹ & John Beamish¹

Superfluidity—liquid flow without friction—is familiar in helium. The first evidence for ‘supersolidity’, its analogue in quantum solids, came from torsional oscillator measurements^{1,2} involving ^4He . At temperatures below 200 mK, the torsional oscillator frequencies increased, suggesting that some of the solid decoupled from the oscillator. This behaviour has been replicated by several groups^{3–7}, but solid ^4He does not respond to pressure differences⁸, and persistent currents and other signatures of superflow have not been seen. Both experiments and theory^{9–14} indicate that defects are involved; these should also affect the solid’s mechanical behaviour. Here we report a measurement of the shear modulus of solid ^4He at low frequencies and strains. We observe large increases below 200 mK, with the same dependence on measurement amplitude, ^3He impurity concentration and annealing as the decoupling seen in the torsional oscillator experiments. We explain this unusual elastic behaviour in terms of a dislocation network that is pinned by ^3He at the lowest temperatures but becomes mobile above 100 mK. The frequency changes in the torsional oscillator experiments appear to be related to the motion of these dislocations, perhaps by disrupting a possible supersolid state.

Although the amount of helium that decouples in different torsional oscillators varies widely, the measurements have many common features. Decoupling occurs below about 200 mK, with a gradual onset accompanied by a dissipation peak at lower temperatures. It decreases at large oscillation amplitudes, which is interpreted in terms of a superflow critical velocity ($v_c \approx 10 \mu\text{m s}^{-1}$). The magnitude of the decoupling is frequency independent, although its onset shifts with frequency⁷. Its amplitude dependence appears to scale with velocity, but depends on the oscillation amplitude during cooling and is hysteretic. A crucial feature of the decoupling is its sensitivity to ^3He . Most measurements used commercial ^4He gas (with ^3He concentration $x_3 \approx 0.3$ p.p.m.), but experiments^{15,16} with isotopically almost pure ^4He (1 p.p.b. ^3He) show a sharper onset at a lower temperature, around 75 mK. Decoupling is usually larger in narrow annuli than in open cylinders⁶ but begins at similar temperatures. Its magnitude also depends on how the solid helium was grown and annealed, indicating that defects are important. Most samples were grown at constant volume under ‘blocked capillary’ conditions—a procedure that involves substantial plastic deformation and is expected to produce a polycrystalline solid with many defects. Theoretical work also suggests that supersolidity does not occur in a perfect crystal^{9,10}, and that grain boundaries¹¹, glassy regions¹² or dislocations^{13,14} are involved. Superflow associated with grain boundaries has been seen in solid ^4He coexisting with liquid¹⁷, but solidification at constant pressure (producing single crystals with fewer defects) still gives significant torsional oscillator decoupling¹⁶. A recently reported heat capacity peak¹⁸ supports the existence of a new phase where decoupling occurs.

We have made a detailed study of the elastic properties of solid ^4He . This required a new method to measure the shear modulus μ at extremely low frequencies and amplitudes, which proved crucial. Embedding piezoelectric transducers in the helium (see Supplementary Fig. 1 and Supplementary Methods) allowed us to measure μ directly at strains as low as $\varepsilon = 2.2 \times 10^{-9}$ (stress $\sigma = 0.03$ Pa). This is two to three orders of magnitude lower than in previous torsional¹⁹, internal friction²⁰ and ultrasonic^{21,22} measurements. We measured μ at frequencies down to 20 Hz, far lower than other experiments. We could also excite and detect acoustic modes of solid helium outside the gap between the transducers, in the surrounding volume. The first such resonance was near 8,000 Hz and had a quality factor $Q \approx 2,000$ at our lowest temperature.

Our essential result is the observation of a large anomalous increase in μ with the same temperature dependence as the decoupling in torsional oscillators. We confirmed this effect by simultaneously measuring the frequency f_r and dissipation $1/Q$ of an acoustic resonance in the cell. A comprehensive study of the anomaly’s dependence on frequency, amplitude and ^3He concentration showed all the same features as the torsional oscillator decoupling, and provided new information about the effects of annealing and stressing the solid. This is, to our knowledge, the first clear observation of directly related phenomena in other properties of solid helium. Our measurements are consistent with the known behaviour of dislocations in solid helium, and lead to a picture of a dislocation network pinned at low temperatures by bound ^3He impurities. Above about 100 mK, ^3He atoms unbind and allow dislocations to move in response to stress, thus weakening the crystal.

Figure 1 shows the temperature dependence of the shear modulus of hexagonal close-packed (h.c.p.) ^4He at pressures of 29.3 and 33.3 bar (with melting temperatures $T_m = 1.77$ and 1.86 K—the first sample passed along the body-centred cubic (b.c.c.)/h.c.p. line during growth). The bottom curve shows μ at 2,000 Hz in the first sample. Below 200 mK, it increases by about 11% ($\Delta\mu \approx 16$ bar): this anomalous stiffening is our central result. The middle curves show this behaviour at three frequencies (2,000, 200 and 20 Hz) in the second sample. The magnitude of the modulus increase, $\Delta\mu$, is similar ($\sim 8\%$) and is nearly independent of frequency, although the transition is sharper at low frequency. The pressure in the cell is constant within 0.2 mbar in this temperature range, which rules out local density changes (for example, freezing of small liquid regions) as the cause of the μ increase, and implies that the bulk modulus does not have a similar anomaly. The upper curve shows a typical non-classical rotational inertia (NCRI) fraction from a torsional oscillator measurement² (at a frequency of 910 Hz). The temperature dependence (onset and shape) is essentially the same as that of the shear modulus anomaly. We observed variations in $\Delta\mu$ of up to a factor of 2 (over a total of 8 samples), similar to the range of NCRI seen in a single torsional oscillator.

¹Department of Physics, University of Alberta, Edmonton, Alberta, T6G 2G7 Canada.

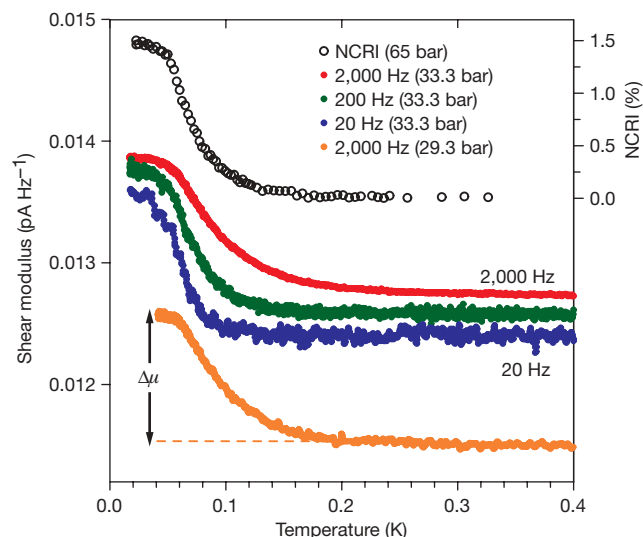


Figure 1 | Shear modulus of solid ^4He at strain $\varepsilon = 2.2 \times 10^{-8}$ as a function of temperature. Shear modulus is given as I/f , where I is the measured current and f is the frequency; data have been offset for clarity. Bottom (orange) curve, μ at 2,000 Hz in a 29.3 bar sample. Middle three curves, μ at 2,000 Hz (red), 200 Hz (green) and 20 Hz (blue) in a 33.3 bar sample. Top curve (open circles, right axes), typical NCRI fraction from a torsional oscillator measurement¹ in a 65 bar sample.

Figure 2 shows μ for the 33.3 bar sample at different strains (calculated from drive voltages). The anomaly $\Delta\mu$ is independent of drive amplitude for strains up to 2.2×10^{-8} then begins to decrease. We observed almost identical behaviour at 200 Hz—the amplitude dependence began at the same drive level, indicating that it scales with stress or strain and not with velocity. The corresponding velocities ($\sim 50 \text{ nm s}^{-1}$ for $\varepsilon = 2.2 \times 10^{-8}$ at 2,000 Hz) are much smaller than the critical velocity $v_c \approx 10 \mu\text{m s}^{-1}$ inferred from torsional oscillator measurements. However, the stress levels (0.3 Pa) are comparable to inertial stresses in torsional oscillators (for example, $\sigma_t \approx 0.15 \text{ Pa}$ at the highest velocity, $520 \mu\text{m s}^{-1}$, in the torsional oscillator of ref. 2). The behaviour is reversible at temperatures above the anomaly and at low amplitudes. When a sample is cooled at high amplitude and the drive is then reduced at low temperature, μ increases (as expected from Fig. 2). However, μ does not decrease when the amplitude is then raised again. Similar hysteretic behaviour is seen in torsional oscillator decoupling⁷. Previous torsional measurements¹⁹ at comparable strains ($\varepsilon = 10^{-7}$) showed no change

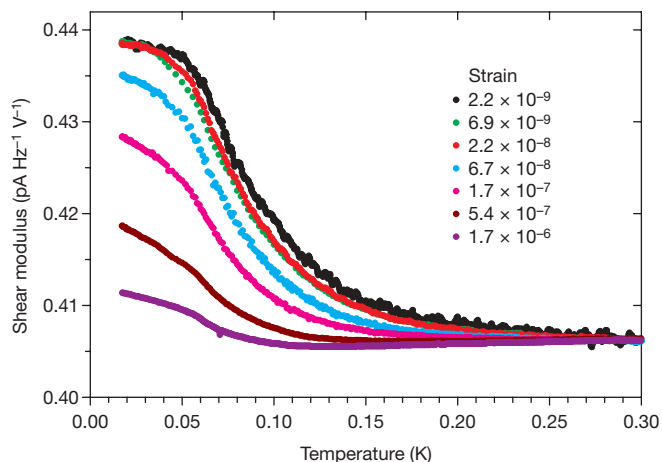


Figure 2 | Shear modulus at 2,000 Hz as a function of peak strain amplitude in the 33.3 bar sample. Shear modulus is given as I/f ; $\varepsilon = 2.2 \times 10^{-9}$ to 1.7×10^{-6} , calculated from transducer drive voltage V ; data for different strains have been shifted to coincide at 300 mK. All data were taken during cooling.

in μ between 0.5 K and 17 mK, an apparent discrepancy with our results.

Figure 3 shows the behaviour of the acoustic resonance near 8 kHz, for the same sample. Comparing the temperature dependence of the resonance frequency f_r to that of the shear modulus μ , it is clear that the two measurements probe the same elastic changes. The changes in f_r are about half as large as for μ ; this is as expected, because f_r varies with sound speed, that is, with the square root of elastic moduli. The corresponding dissipation $1/Q$ is largest near 150 mK, where f_r is changing rapidly (the dissipation peak occurs at lower temperatures in most torsional oscillator measurements, but this may be due to their lower frequencies, 185–1,500 Hz). In a simple oscillator, the maximum dissipation $\Delta(1/Q)$ should equal the frequency shift $\Delta f_r/f_r$, but in our case it is a factor of about 3 smaller. Similar differences in torsional oscillator measurements have been ascribed²³ to sample inhomogeneity.

A striking feature of torsional oscillator experiments is their sensitivity to ^3He . We grew samples from isotopically almost pure ^4He (1 p.p.b. ^3He —the same gas used in torsional oscillator measurements¹⁵) and from gases with intermediate concentrations of ^3He made by mixing with commercial ^4He (0.3 p.p.m. ^3He). We compare their behaviour in Fig. 4. Changes have been scaled by $\Delta\mu$ at 18 mK to compare temperature dependences. The anomaly shifts to lower temperatures as the ^3He concentration decreases. We also show similarly scaled decoupling data from torsional oscillator experiments^{2,16} on 1 p.p.b. and 0.3 p.p.m. ^3He samples: the onset temperatures and shapes of the curves agree very well (within the sample to sample variations in torsional oscillator measurements).

To understand the role of defects, we annealed the 33.3 bar sample at 160 mK below its melting point T_m for 15 h. This reduced $\Delta\mu$ from 9.8% to 7.7%, but it was the high-temperature behaviour that changed. The low-temperature values of μ and f_r were almost unaffected (for example, at 18 mK, f_r increased by only 0.1%) and appear to reflect the intrinsic shear modulus, so the effect of annealing is to change μ at higher temperatures. We also applied large acoustic stresses ($\geq 700 \text{ Pa}$) to the annealed crystal in an attempt to create additional defects. Again, the values of μ and f_r changed at high

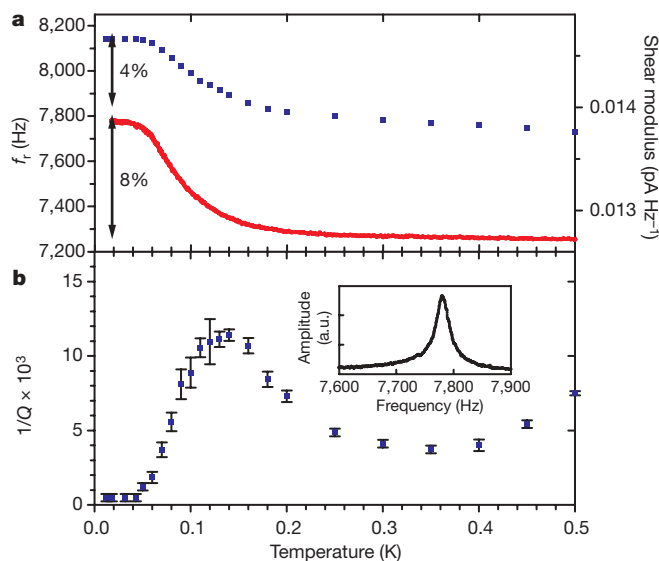


Figure 3 | Temperature dependence of acoustic resonance peak frequency f_r and dissipation $1/Q$ for the 33.3 bar sample. **a**, Resonance frequency (upper blue points, left axis) and, for comparison, shear modulus at 2,000 Hz (lower red curve, right axis). Vertical arrows are for scale, showing 4% and 8% changes in f_r and μ , respectively. **b**, Dissipation (calculated from the full-width at half-maximum of the peaks) corresponding to resonance data in **a**. Inset, typical resonance peak at 300 mK with a $Q \approx 250$. Error bars reflect the maximum and minimum peak widths consistent with the statistical noise in the voltage amplitude.

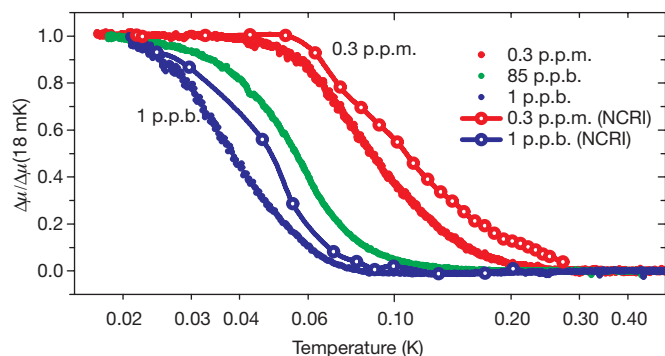


Figure 4 | Shear modulus anomaly in solid ^4He for a range of ^3He impurity concentrations. Concentrations are 1 p.p.b. (33.3 bar, blue symbols), 85 p.p.b. (32.9 bar, green symbols) and 0.3 p.p.m. ^3He (33.3 bar, red symbols). Changes $\Delta\mu$ have been scaled by the values at the lowest temperature (18 mK) in order to compare temperature dependences. Open circles with lines are similarly scaled NCI data from torsional oscillator measurements on 1 p.p.b. ^3He (31 bar, blue symbols¹⁴) and 0.3 p.p.m. ^3He (26 bar, red symbols¹) samples.

temperature but not at low temperature. Warming above ~ 0.6 K undoes these effects, indicating that defects introduced by stressing the crystal are only stable at low temperatures. Ultrasonic measurements²⁴ on b.c.c. and h.c.p. ^3He showed similar effects of large stresses.

The modulus changes we see are much larger than expected in defect-free crystals. Dislocations (one-dimensional defects created during crystal growth or deformation) can dramatically affect elastic properties and lead to unusual behaviour in quantum crystals²⁵. They form three-dimensional networks, pinned where they intersect, and characterized by their density Λ (total length per unit volume), Burgers vector \mathbf{b} , and length between intersections L_N . They can also be pinned, less strongly, by impurities with the pinning length L_p determined by the impurity concentration x , binding energy E_B , and the temperature.

Dislocations move in response to stress, producing a strain²⁶. At low frequencies this reduces the shear modulus by $\Delta\mu/\mu = -CR\Lambda L^2$ (see Supplementary Discussion), where C is a constant (≤ 0.5) that depends on the distribution of lengths, R is an orientation factor (~ 0.2) and L is the effective dislocation loop length. In annealed crystals with well-defined networks, ΛL_N^2 is a geometric constant (for example, 3 for a cubic network), so dislocations can produce a frequency-independent reduction of the shear modulus as large as 30%. Impurities can immobilize dislocations and restore the crystal's intrinsic modulus. This pinning becomes important when L_p is comparable to L_N , which occurs at a temperature T_p that decreases with impurity concentration: $T_p \approx -(E_B/k_B)[\ln(xL_N/a)]^{-1}$, where k_B is Boltzmann's constant and a is the interatomic spacing (~ 0.35 nm).

Our results are consistent with a picture of a network of dislocations pinned at low temperature by ^3He impurities, using dislocation parameters determined in earlier experiments on h.c.p. ^4He . Low-frequency torsional measurements¹⁹ on helium gave a ^3He binding energy $E_B/k_B \approx 0.6$ K and $R\Lambda L^2 \approx 1$. Ultrasonic measurements^{21,22} on single crystals gave $L_N \approx 5$ μm . Using these values, $T_p \approx 110$ mK for $x_3 = 0.3$ p.p.m., decreasing to 54 mK for $x_3 = 1$ p.p.b., close to the temperatures where μ increases. Large stresses can tear dislocations away from ^3He pinning sites and reduce the shear modulus²⁷. The critical stress for this breakaway can be estimated²⁷ as 4 Pa for $L_p \approx 5$ μm . This corresponds to $\epsilon \approx 3 \times 10^{-7}$, a strain where we see strong amplitude dependence. Annealing should reduce the dislocation density Λ (and large stresses can increase Λ), but $\Delta\mu$ depends on the combination ΛL^2 , which does not necessarily decrease when dislocations disappear. The low-temperature behaviour is unaffected by annealing or stressing; this is as expected because dislocations are then pinned by impurities and do not affect μ . When the temperature

is raised, ^3He impurities 'boil off' the dislocations, allowing them to move and reducing μ .

Is our shear modulus anomaly directly related to the frequency shifts and dissipation in torsional oscillator experiments? Although they measure very different properties (shear modulus and sound speed versus moment of inertia and density), the two sets of measurements share all essential features. The anomalous behaviour has the same temperature dependence and the transitions are accompanied by similar dissipation peaks. Both are strongly amplitude dependent (starting at comparable stress levels) and have similar amplitude-dependent hysteresis at low temperatures. In both types of measurements, the anomaly's magnitude is frequency independent, but its onset is broadened and shifts to higher temperature with increasing frequency. Minute ^3He concentrations have the same dramatic effect on the onset temperature, and annealing changes the size of both anomalies. Given these remarkable similarities, the two sets of effects must be closely related. The obvious question is 'how?'

One possibility is that the modulus increase stiffens the torsional oscillator, increasing its frequency and mimicking mass decoupling. Interpreting a torsional oscillator frequency as a direct measure of mass assumes that the oscillator head is infinitely stiff and that the solid helium moves rigidly with its walls, neither of which is exactly true. However, an increase in the helium's shear modulus would improve its coupling to the torsional oscillator and thus decrease its frequency (that is, the opposite of the observed behaviour). An increase in μ could raise an oscillator's frequency by increasing the stiffness of its head. However, the head itself is much stiffer than the torsion rod and the helium's contribution to the head's stiffness is small (it has moduli about 3,000 times smaller than copper), so modulus changes in the helium should produce very small frequency changes. Estimates for typical oscillators indicate that this effect is several orders of magnitude too small to explain the observed decoupling. Also, blocking the flow path in a torsional oscillator annulus would barely change the helium's contribution to its stiffness, yet nearly eliminated the decoupling². Our observations do not provide an obvious mechanical, non-supersolid explanation²⁸ of the frequency changes in torsional oscillators.

Our μ anomaly and the decoupling seen in torsional oscillator measurements could both be fundamental properties of a supersolid state²⁹, in which case it would be natural for them to have a common dependence on temperature, ^3He , and so on. This could be the case if, for example, supersolidity occurs along the cores¹³ or strain fields¹⁴ of dislocation networks. Alternatively, the μ anomaly could, as we propose, be due to dislocations becoming mobile, which in turn could affect a supersolid response. For example, vortices³⁰ could be pinned by stationary dislocations but could introduce dissipation and destroy the supersolidity when dislocations begin to move above 100 mK. The decoupling seen in porous media¹ remains a puzzle, as it is hard to imagine dislocations existing, let alone moving, in the 7 nm pores of Vycor glass. The precise connection between our elastic measurements and decoupling of solid helium from torsional oscillators is not certain, but it is clear that the two are closely related and that models of supersolidity should consider the effects of moving dislocations.

Received 21 August; accepted 2 October 2007.

- Kim, E. & Chan, M. H. W. Probable observation of a supersolid helium phase. *Nature* **427**, 225–227 (2004).
- Kim, E. & Chan, M. H. W. Observation of superflow in solid helium. *Science* **305**, 1941–1944 (2004).
- Kondo, M., Takada, S., Shihayama, Y. & Shirahama, K. Observation of non-classical rotational inertia in bulk solid ^4He . *J. Low Temp. Phys.* **148**, 695–699 (2007).
- Penzev, A., Yasuta, Y. & Kubota, M. Annealing effect for supersolid fraction in ^4He . *J. Low Temp. Phys.* **148**, 677–681 (2007).
- Rittner, A. S. C. & Reppy, J. D. Observation of classical rotational inertia and nonclassical supersolid signals in solid ^4He below 250 mK. *Phys. Rev. Lett.* **97**, 165301 (2006).

6. Rittner, A. S. C. & Reppy, J. D. Disorder and the supersolid state of solid ^4He . *Phys. Rev. Lett.* **98**, 175302 (2007).
7. Aoki, Y., Graves, J. C. & Kojima, H. Oscillation frequency dependence of nonclassical rotation inertia of solid ^4He . *Phys. Rev. Lett.* **99**, 015301 (2007).
8. Day, J. & Beamish, J. Pressure-driven flow of solid helium. *Phys. Rev. Lett.* **96**, 105304 (2006).
9. Ceperley, D. M. & Bernu, B. Ring exchanges and the supersolid phase of ^4He . *Phys. Rev. Lett.* **93**, 155303 (2004).
10. Prokof'ev, N. & Svistunov, B. Supersolid state of matter. *Phys. Rev. Lett.* **94**, 155302 (2005).
11. Pollet, L. *et al.* Superfluidity of grain boundaries in solid ^4He . *Phys. Rev. Lett.* **98**, 135301 (2007).
12. Boninsegni, M., Prokof'ev, N. & Svistunov, B. Superglass phase of ^4He . *Phys. Rev. Lett.* **96**, 105301 (2006).
13. Boninsegni, M. *et al.* Luttinger liquid in the core of a screw dislocation in helium-4. *Phys. Rev. Lett.* **99**, 035301 (2007).
14. Toner, J. Quenched disorder enhanced supersolid ordering. *arXiv:0707.3842* (2007).
15. Kim, E., Xia, J. S., West, J. T., Lin, X. & Chan, M. H. W. Effect of ^3He impurity on the supersolid transition of ^4He . *Bull. Am. Phys. Soc.* **52**, 610 (2007).
16. Clark, A. C., West, J. T. & Chan, M. H. W. Non-classical rotational inertia in helium crystals. Preprint at (<http://arxiv.org/abs/0706.0906>) (2007).
17. Sasaki, S., Ishiguro, R., Caupin, F., Maris, H. J. & Balibar, S. Superfluidity of grain boundaries and supersolid behavior. *Science* **313**, 1098–1100 (2006).
18. Lin, X., Clark, A. C. & Chan, M. H. W. Probable heat capacity signature of the supersolid transition. *Nature* **449**, 1025–1028 (2007).
19. Paalanen, M. A., Bishop, D. J. & Dail, H. W. Dislocation motion in hcp ^4He . *Phys. Rev. Lett.* **46**, 664–667 (1981).
20. Tsymbalenko, V. L. Measurement of internal friction in solid ^4He . *Sov. Phys. JETP* **47**, 787–792 (1978).
21. Iwasa, I., Araki, K. & Suzuki, H. Temperature and frequency dependence of the sound velocity in hcp ^4He crystals. *J. Phys. Soc. Jpn.* **46**, 1119–1126 (1979).
22. Beamish, J. R. & Franck, J. P. Sound propagation at frequencies from 3 to 21 MHz in hcp and bcc ^3He and its interaction with dislocations. *Phys. Rev. B* **26**, 6104–6113 (1982).
23. Huse, D. A. & Khandker, Z. U. Dissipation peak as an indicator of sample inhomogeneity in solid ^4He oscillator experiments. Preprint at (<http://arxiv.org/abs/cond-mat/0702243>) (2007).
24. Beamish, J. R. & Franck, J. P. Pinning of dislocations in hcp and bcc ^3He by stress waves and by ^4He impurities. *Phys. Rev. B* **28**, 1419–1432 (1983).
25. De Gennes, P.-G. Quantum dynamics of a single dislocation. *C.R. Phys.* **7**, 561–566 (2006).
26. Granato, A. & Lucke, K. Theory of mechanical damping due to dislocations. *J. Appl. Phys.* **27**, 583–593 (1956).
27. Iwasa, I. & Suzuki, H. Sound velocity and attenuation in hcp ^4He crystals containing ^3He impurities. *J. Phys. Soc. Jpn.* **49**, 1722–1730 (1980).
28. Nussinov, Z. *et al.* Origin of the decrease in the torsional-oscillator period of solid ^4He . *Phys. Rev. B* **76**, 014530 (2007).
29. Dorsey, A. T., Goldbart, P. M. & Toner, J. Squeezing superfluid from a stone: coupling superfluidity and elasticity in a supersolid. *Phys. Rev. Lett.* **96**, 055301 (2006).
30. Anderson, P. W. Two new vortex fluids. *Nature Physics* **3**, 160–162 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Natural Sciences and Engineering Research Council of Canada and the University of Alberta for support of this research; we also thank M. H. W. Chan for providing the torsional oscillator data of Figs 1 and 4.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.B. (beamish@phys.ualberta.ca).

Controlling cavity reflectivity with a single quantum dot

Dirk Englund^{1*}, Andrei Faraon^{1*}, Ilya Fushman^{1*}, Nick Stoltz², Pierre Petroff² & Jelena Vučković¹

Solid-state cavity quantum electrodynamics (QED) systems offer a robust and scalable platform for quantum optics experiments and the development of quantum information processing devices. In particular, systems based on photonic crystal nanocavities and semiconductor quantum dots have seen rapid progress. Recent experiments have allowed the observation of weak¹ and strong coupling^{2,3} regimes of interaction between the photonic crystal cavity and a single quantum dot in photoluminescence. In the weak coupling regime¹, the quantum dot radiative lifetime is modified; in the strong coupling regime³, the coupled quantum dot also modifies the cavity spectrum. Several proposals for scalable quantum information networks and quantum computation rely on direct probing of the cavity–quantum dot coupling, by means of resonant light scattering from strongly or weakly coupled quantum dots^{4–9}. Such experiments have recently been performed in atomic systems^{10–12} and superconducting circuit QED systems¹³, but not in solid-state quantum dot–cavity QED systems. Here we present experimental evidence that this interaction can be probed in solid-state systems, and show that, as expected from theory, the quantum dot strongly modifies the cavity transmission and reflection spectra. We show that when the quantum dot is coupled to the cavity, photons that are resonant with its transition are prohibited from entering the cavity. We observe this effect as the quantum dot is tuned through the cavity and the coupling strength between them changes. At high intensity of the probe beam, we observe rapid saturation of the transmission dip. These measurements provide both a method for probing the cavity–quantum dot system and a step towards the realization of quantum devices based on coherent light scattering and large optical nonlinearities from quantum dots in photonic crystal cavities.

In the experiment, a narrow-bandwidth laser beam is scanned through the resonance of a GaAs photonic crystal cavity (Fig. 1c). The cavity contains a strongly coupled InAs quantum dot that splits its spectrum into two polariton states and causes the cavity transmission to vanish at the quantum dot frequency¹⁴. A linear three-hole defect in the photonic crystal forms the cavity¹⁵ with a resonant mode at $\lambda = 926$ nm and measured quality factor $Q = 1.0 \times 10^4$ (corresponding to a cavity linewidth of $\Delta\lambda_{\text{cav}} \approx 0.10$ nm). We observe a polariton splitting of 0.05 nm. The photonic crystal was fabricated on a quantum dot wafer grown by molecular beam epitaxy, as described in Methods.

The principle of the measurement is explained in Fig. 1b. It is difficult to observe the cavity spectrum directly, because only a small fraction of the incident light couples to the photonic crystal cavity owing to poor mode matching between the gaussian probe beam and the cavity mode. For that reason, the signal reflected by the cavity is monitored in cross-polarization. This is analogous to observing transmission through a polarizing cavity inserted between two

crossed polarizers. A GaAs/AlAs distributed Bragg reflector underneath the photonic crystal membrane effectively creates a single-sided cavity system and enhances the collection efficiency of the probe beam. The horizontal $|H\rangle$ component of the scattered probe beam then carries the cavity reflectivity R , as given by equation (2).

Reflectivity is measured by scanning the narrow-linewidth probe laser beam through the cavity resonance (Fig. 1a, b). In this way, we greatly exceed the 0.03 nm resolution of the spectrometer in order to sample the narrow spectral features of the system (that is, 0.05 nm Rabi splitting). To avoid difficulties related to laser stability and power normalization, we keep the laser wavelength fixed and instead scan the cavity and quantum dot using our recently developed local temperature-tuning technique¹⁶. The technique uses a laser beam to heat the suspended structure depicted in Fig. 1c, which is composed of a photonic crystal cavity and a heating pad. The structure was fabricated by electron beam lithography and reactive ion etching. The pad is coated with a Cr/Au metal layer to increase absorption of the 905 nm heating laser, which is tuned to this wavelength to minimize the carrier excitation in GaAs and thus reduce background photoluminescence. The sample is maintained at an average temperature of 27 K and probed using the confocal microscope set-up in Fig. 1a. The reflectivity signal from a different cavity without coupled quantum dots is shown in Fig. 1d. Here, the cavity resonance is swept through the tunable probe laser line using the local heating technique. A half-wave plate in front of the sample corrects for non-optimal orientation of the cavity and maximizes its visibility in the reflected signal (Fig. 1a). We verified that the visibility vanishes when the probe polarization is orthogonal or parallel to the cavity polarization. We obtain a cavity signal-to-background ratio of unity, which together with the imperfect extinction ratio of the polarizing beam splitter, lets us estimate that the coupling efficiency into the cavity mode is 1–2%. A more detailed explanation of how the measurement was performed is presented in Methods.

We first characterize the quantum dot–photonic crystal cavity system by photoluminescence when pumped with a continuous-wave laser beam at 780 nm, above the GaAs bandgap (incident power ~ 20 nW before the objective). For low excitation powers, the quantum dot photoluminescence increases linearly with pump power, indicating a single exciton line. As the quantum dot is temperature-tuned through the cavity, clear anticrossing between the quantum dot and the cavity lines is observed: the quantum dot splits the cavity spectrum into two polariton peaks (with frequencies ω_{\pm}) when it becomes resonant with the cavity (Fig. 2c). This splitting is described by

$$\omega_{\pm} = \frac{\omega_c + \omega_d}{2} - i\frac{\kappa + \gamma}{2} \pm \sqrt{g^2 + \frac{1}{4}(\delta - i(\kappa - \gamma))^2} \quad (1)$$

where ω_c denotes the cavity frequency, ω_d the quantum dot frequency, $\delta = \omega_d - \omega_c$ the quantum dot–cavity detuning, cavity field decay rate $\kappa/2\pi = 16$ GHz (linewidth 0.1 nm), Rabi frequency $g/2\pi = 8$ GHz

¹Ginzton Laboratory, Stanford University, Stanford, California 94305, USA. ²Department of Electrical and Computer Engineering, University of California, Santa Barbara, California 93106, USA.

*These authors contributed equally to this work.

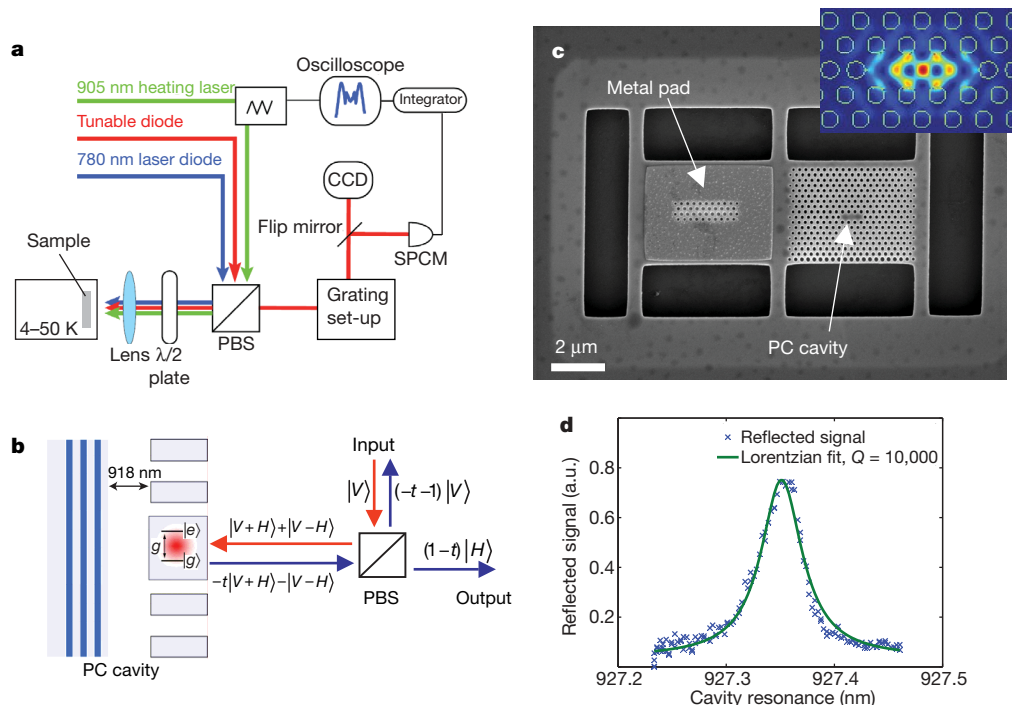


Figure 1 | Experiment set-up. **a**, Confocal microscope set-up. A 780 nm laser diode excites photoluminescence, while a 905 nm modulated Ti:sapphire laser locally heats the sample to tune cavity and quantum dot¹⁶. The reflectivity is measured with a narrow-band tunable diode laser (focal spot diameter $\sim 1 \mu\text{m}$ for all beams). A grating set-up monitors the photoluminescence and filters the reflectivity signal from background noise. The filtered reflected signal is detected by a single photon counting module (SPCM). **b**, Principle of the reflectivity measurement off a photonic crystal (PC) cavity. A vertical ($|V\rangle$ -polarized) probe laser is directed onto the linearly polarized cavity oriented at 45° ($|V+H\rangle$). Owing to interaction with

the cavity, the $|V+H\rangle$ component of the probe beam is reflected with a frequency-dependent coefficient $-t(\omega)$. The $|V-H\rangle$ component reflects directly with a π phase shift. The polarizing beam splitter (PBS) passes $|H\rangle$, giving a signal that is proportional to $|1-t|^2$ on the detector (see equation (2)). **c**, Suspended structure composed of a heating pad and a photonic crystal cavity. The heating laser incident on the metal pad controls local temperature¹⁶. Inset, simulated electric field intensity of photonic crystal cavity. **d**, Reflectivity spectrum obtained by tuning an empty cavity (no coupled quantum dot) through the probe laser, indicating $Q = 1.0 \times 10^4$.

(from Rabi splitting of $2g$ corresponding to 0.05 nm), and the dipole decay rate without the cavity $g/2\pi \approx 0.1 \text{ GHz}$. As $g \approx \kappa/2$, the cavity–quantum dot system operates at the onset of strong coupling¹⁴, as was also the case for other quantum dot–photonic crystal cavity QED experiments done in photoluminescence^{2,3}.

To accurately interpret the photoluminescence and reflectivity data, we need to know the frequency of the cavity and strongly

coupled quantum dot. Direct tracking of the latter is difficult because of its modified spectrum when coupled to the cavity, and because it rapidly decreases in intensity as it exits the cavity (Fig. 2b). This problem is solved by instead tracking a nearby quantum dot that precisely follows, at a fixed offset, the strongly coupled quantum dot's trajectory (Fig. 2a). Based on this, the strongly coupled quantum dot wavelength is shown in the inset of Fig. 2b, together with that of the

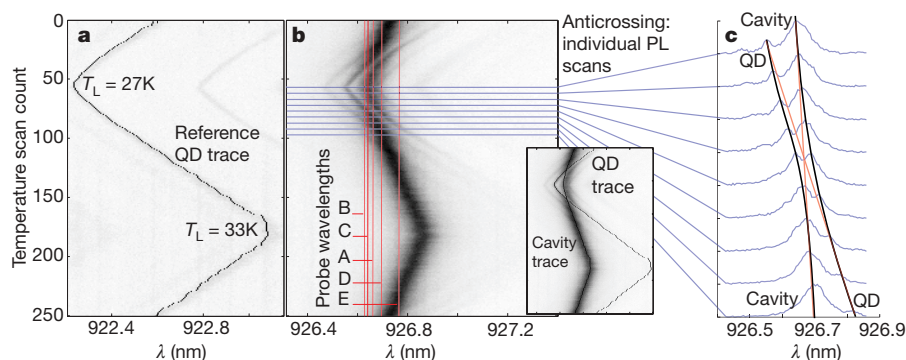


Figure 2 | Photoluminescence of a single quantum dot tuned through strong coupling to a photonic crystal cavity. The dot is excited using an above-band pump beam (780 nm wavelength, with 20 nW power incident on the sample surface). Tuning of the quantum dot through the cavity resonance is achieved following our earlier work¹⁶, with a heating beam intensity-modulated between $6 \mu\text{W}$ and $300 \mu\text{W}$ to change local temperature T_L from 27 K to 33 K . **a**, A reference quantum dot (QD) is used for tracing the wavelength of the strongly coupled quantum dot, as dots that are closely spaced in wavelength exhibit identical temperature tuning behaviour. The heating beam power is modulated with a triangular pattern and shifts the

quantum dot nearly linearly. **b**, Photoluminescence (PL) emission shows the strongly coupled quantum dot tuned in and out of resonance with a photonic crystal cavity ($Q \approx 1.0 \times 10^4$). In the reflectivity measurements, the above-band pump is switched off and the cavity/quantum dot system probed at different detunings of the reflected laser beam from the point of anticrossing (lines A–E). Inset, quantum dot and cavity traces. **c**, Individual photoluminescence cross-sections show anticrossing between quantum dot and cavity, with measured Rabi splitting of 0.05 nm (corresponding to $2g$, where the coupling strength $g/2\pi = 8 \text{ GHz}$). As a guide the eye, we show the wavelengths of the uncoupled quantum dot and cavity (red line).

cavity, which shifts at a rate equal to 0.28 of the rate of the quantum dot shift.

The reflectivity of the quantum dot–cavity system is probed at five different spectral detunings $\Delta\lambda = \lambda - \lambda_0$ of the probe laser from λ_0 , the anticrossing point of quantum dot and cavity (inset Fig. 3a). The incident power is in the weak excitation limit at 3 nW (measured before the objective lens), corresponding to less than one photon inside the cavity per cavity lifetime, as required for probing the vacuum Rabi splitting. For each reflectivity scan, a corresponding photoluminescence scan is obtained to track quantum dot and cavity wavelengths. Figure 3 plots the reflectivity signal as a function of temperature scan. In this data set, the temperature tuning is used to sweep the quantum dot and cavity back and forth through the probe laser. These data form the central measurement of this paper: as the single quantum dot sweeps across the cavity, it strongly modifies the reflected intensity. Instead of observing a lorentzian-shaped cavity spectrum (Fig. 1d), a drop in the reflected signal is observed at the quantum dot wavelength, as expected in the strong coupling regime. From a quantum mechanical perspective, when the quantum dot is on resonance with the cavity and strongly coupled to it, the quantum dot–cavity system does not have an energy eigenstate at the bare quantum dot resonance, and photons resonant with the quantum dot cannot be coupled into the cavity (Fig. 1b).

The reflected signal from the described cavity is derived following refs 6 and 8. The spectrum R of the reflected probe signal after the polarizing beam splitter is then given by

$$R = \eta \left| \frac{\kappa}{i(\omega_c - \omega) + \kappa + \frac{g^2}{i(\omega_d - \omega) + \gamma}} \right|^2 \quad (2)$$

where η accounts for the efficiency of coupling to, and collecting from, the cavity. We fitted this relation to the observed spectrum, using the above-mentioned cavity–quantum dot parameters, together with the tracked quantum dot and cavity wavelengths shown at the bottom of Fig. 3a. The experimental data in the top panel of Fig. 3a show smoother features than the plot of equation (2) based on tracked quantum dot and cavity lines (dashed line). We attribute this difference to spectral fluctuations in the quantum dot and cavity that are below the resolution limit of the spectrometer, but that are greater than the linewidth of the probe beam. These fluctuations arise from instabilities in the power of the heating laser of $\sim 0.7\%$. When thermal fluctuation in the quantum dot–cavity wavelength is taken into account as a gaussian broadening with full-width

at half-maximum of 0.005 nm, the theoretical model matches the data (black fits).

The fits yield values for coupling strength g and cavity Q that agree with photoluminescence measurements in above-band pumping. The reflectivity data for the other probe wavelengths (Fig. 3b) capture the quantum dot at various detunings from the cavity–quantum dot intersection ranging from $-1.2g$ (-0.03 nm) to $4.5g$ (0.11 nm). The reflected probe drops towards zero precisely where the quantum dot crosses its wavelength, and the depth and shape of the drop changes with cavity detuning as predicted by theory. We note that an alternative model of an absorbing quantum dot¹⁷ inside the cavity does not fit the reflectivity data, and predicts a cavity spectral linewidth that does not agree with the measured value. These measurements also point to one of the advantages of the solid-state cavity QED system: it is possible to capture the spatially fixed quantum dot in various states of detuning and at constant coupling to the cavity, whereas atomic systems are complicated by moving emitters.

In Fig. 4, we explore the nonlinear behaviour of another strongly coupled quantum dot–photonic crystal cavity system as a function of power P_{in} of the probe laser beam. This system shows the same coupling strength as the first, with $g/2\pi = 8$ GHz and $Q = 10^4$, and is probed here when the quantum dot is detuned by $\Delta\lambda = -0.012$ nm (corresponding to $-g/2$) from the anticrossing. P_{in} is increased from the low-excitation limit at 5 nW before the objective (corresponding to an average cavity photon number $\langle n_{\text{cav}} \rangle \approx 0.003$ in a cavity without a quantum dot) to the high-excitation regime with $P_{\text{in}} \approx 12$ μ W (corresponding $\langle n_{\text{cav}} \rangle \approx 7.3$). Here, $\langle n_{\text{cav}} \rangle$ is estimated as $\eta P_{\text{in}} / 2\kappa\hbar\omega_c$, where $\eta \approx 1.8\%$ is the coupling efficiency into the cavity at this wavelength. Figure 4a shows the quantum-dot-induced reflectivity dip vanishing as P_{in} is increased by roughly three orders of magnitude. We modelled the saturation behaviour by a steady-state solution of the quantum master equation following ref. 18, using the above-mentioned measured system parameters. The cavity mode is represented by a number state basis truncated to $n = 100$ and driven by a coherent electric field with varying amplitude E . Figure 4a also plots the calculated normalized reflected intensity as a function of the cavity and quantum dot tuning with temperature (solid line). We see very good agreement when the solution is convolved with the gaussian filter accounting for spectral fluctuations arising from heating noise, as explained above. The full data are summarized in Fig. 4b, where we plot the reflectivity R at the quantum dot detuning $\Delta\lambda = -0.012$ nm, normalized by the reflectivity value R_0 for an empty cavity at the same wavelength as the probe laser (that is, for $g \rightarrow 0$). Our results agree with

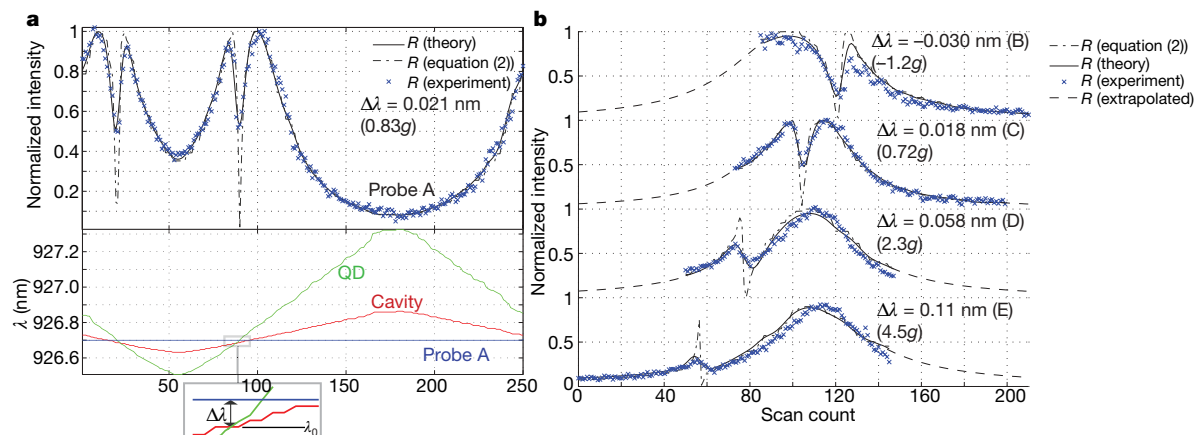


Figure 3 | Quantum dot-controlled cavity reflectivity at different probe wavelengths A–E, as indicated in Fig. 2b. **a**, Reflectivity spectrum of probe laser as function of quantum dot and cavity detunings, as determined from corresponding photoluminescence spectra (Fig. 2). The probe laser is detuned by $\Delta\lambda = 0.021$ nm (corresponding to $\Delta\lambda = 0.83g$) from the anticrossing point λ_0 between quantum dot and cavity (see inset). Ideal

theoretical plots are calculated from equation (2). Also shown are theoretical plots that take into account a jitter (~ 0.005 nm) of cavity and quantum dot wavelength resulting from the heating laser power fluctuation. **b**, Probe laser at various detunings $\Delta\lambda$ from the anticrossing point samples different quantum dot–cavity detunings. Incomplete scans result from the limited range of temperature tuning.

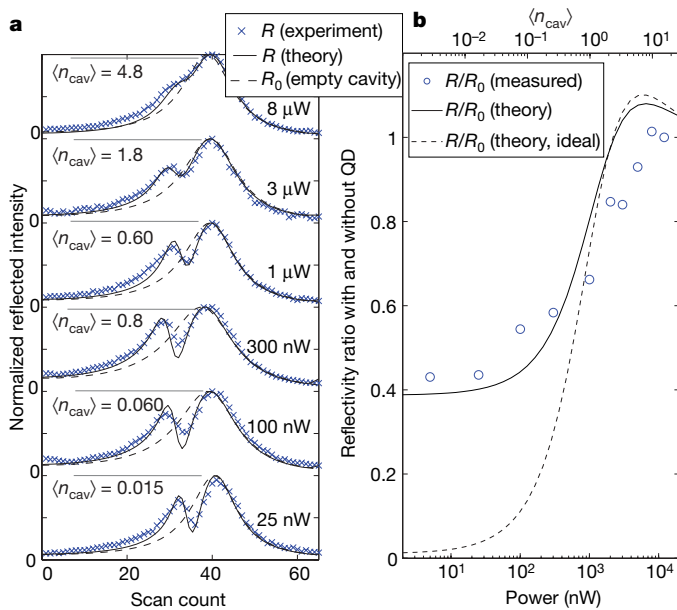


Figure 4 | Quantum-dot-controlled cavity reflectivity versus probe beam power for probe laser detuning of $\Delta\lambda = -0.012$ nm from the anticrossing point. **a**, Reflectivity scans at increasing probe power (measured before the objective), ranging from low-excitation to saturation regimes. The measured reflectivity is fitted by a numerical solution to the full master equation. Solutions are convolved with a gaussian filter with full-width at half-maximum of 0.005 nm to account for thermal fluctuations (solid curves). The scale for the calculated mean photon number $\langle n_{\text{cav}} \rangle$ is also indicated for each scan. Also plotted is the expected reflectivity R_0 when the quantum dot is removed (dashed curve). **b**, Reflectivity at $\Delta\lambda = -0.012$, normalized by empty-cavity reflectivity at the same wavelength, as a function of probe laser power. Saturation begins near 1 μW of input power, corresponding to $\langle n_{\text{cav}} \rangle \approx 1/2$. The dashed curve shows the reflectivity ratio if no thermal fluctuations were present. At large power, both curves tend to unity as the quantum dot–cavity spectrum approaches the lorentzian shape of the empty cavity.

the theoretical model (solid curve) and previous measurements in atomic systems¹⁹. Owing to the spectral fluctuations, the reflectivity does not approach zero at low power, as it would in the ideal system (dashed curve). Saturation begins at ~ 1 μW of incident power (measured before the objective), corresponding to $\langle n_{\text{cav}} \rangle \approx 1/2$. Taking into account the coupling efficiency η , this implies a saturation power inside the cavity of only ~ 20 nW, in agreement with previous predictions for giant optical nonlinearity in a microcavity²⁰. We furthermore verified that the quantum-dot-induced reflectivity dip vanishes controllably when excitons are (incoherently) generated by excitation with an above-GaAs-bandgap laser beam.

In conclusion, we have experimentally demonstrated that a single quantum dot can be used to dramatically alter the reflectivity spectrum of an optical cavity. In the low-excitation regime (intracavity photon number $\langle n_{\text{cav}} \rangle \ll 1$), we observe a quantum-dot-induced change in reflectivity to 40% and find very good agreement with theory. The remaining signal is limited by measurement noise (that is, quantum dot and cavity wavelength fluctuations resulting from power instabilities of the heating laser), and should vanish with improved experimental stability. As the resonant beam intensity is increased, we observe saturation of the quantum-dot-induced dip at ~ 20 nW of cavity-coupled power (photon number $\langle n_{\text{cav}} \rangle \approx 1/2$), closely matched by theory. Our measurements rely on a novel quantum dot–cavity tuning and cross-polarized reflectivity method that permits resolution of ~ 0.005 nm (full-width at half-maximum) and high cavity–quantum dot visibility. The photonic crystal architecture is ideally suited for extending this system to greater numbers of quantum dots and cavities interconnected into a quantum network²¹. Such an on-chip approach greatly increases the coupling

efficiency to and from the cavity²², and our recent circuits should allow efficiencies exceeding 50% while ensuring a cavity $Q > 10^4$. The demonstration of quantum-dot-controlled cavity reflectivity has far-reaching implications for quantum information processing in solid-state systems, as it opens the door to high-fidelity controlled phase gates⁶, single photon detection¹², coherent transfer of quantum dot state to photon state⁴, and quantum repeaters using non-destructive Bell measurements with the addition of a third long-lived quantum dot level⁸. The observed giant optical nonlinearity has promising applications for generating non-classical squeezed states of light^{10,23}, non-destructive photon number state measurements²⁴, and optical signal processing.

METHODS SUMMARY

Reflectivity measurement. A cavity with coupled quantum dot showing polariton anti-crossing was first identified in photoluminescence, using above-band excitation at 780 nm. The temperature of the cryostat and the power of the heating laser were controlled so the quantum dot periodically swept through the cavity resonance. Then the tunable diode laser used for reflectivity measurements was set to the desired wavelength using the spectrometer. After spectral alignment, the 780 nm laser was turned off, and the reflectivity signal was sent to a photodetector and optimized on an oscilloscope. Once optimized, the output was switched to the spectrometer CCD and the reflectivity signal was recorded with the spectrometer taking successive spectra at 0.2-s-long integration, while the heating laser power (and subsequently quantum dot and cavity wavelength) was modulated at 10 mHz. This scanning speed is slow enough to resolve the relevant features, as seen by the number of data points sampling the quantum-dot-induced dips in Fig. 3.

Quantum dot wafer. The photonic crystal was fabricated on a quantum dot wafer grown by molecular beam epitaxy on a Si n-doped GaAs(100) substrate with a 0.1 μm buffer layer, and a 10-period distributed Bragg reflector consisting of quarter-wave AlAs/GaAs layers to improve collection efficiency into the lens²⁵. The distributed Bragg reflector is separated by a 918 nm sacrificial layer of Al_{0.8}Ga_{0.2}As from the 150-nm GaAs membrane that contains a central layer of self-assembled InGaAs/GaAs quantum dots. The quantum dot density varies throughout the wafer, but in this experiment, we used the low-density area with ~ 100 quantum dots per μm².

Received 5 June; accepted 4 September 2007.

- Englund, D. et al. Controlling the spontaneous emission rate of single quantum dots in a two-dimensional photonic crystal. *Phys. Rev. Lett.* **95**, 013904 (2005).
- Yoshie, T. et al. Vacuum Rabi splitting with a single quantum dot in a photonic crystal nanocavity. *Nature* **432**, 200–203 (2004).
- Hennessy, K. et al. Quantum nature of a strongly coupled single quantum dot-cavity system. *Nature* **445**, 896–899 (2007).
- Cirac, J. I., Zoller, P., Kimble, H. J. & Mabuchi, H. Quantum state transfer and entanglement distribution among distant nodes in a quantum network. *Phys. Rev. Lett.* **78**, 3221–3224 (1997).
- Imamoglu, A. et al. Quantum information processing using quantum dot spins and cavity QED. *Phys. Rev. Lett.* **83**, 4204–4207 (1999).
- Duan, L. M. & Kimble, H. J. Scalable photonic quantum computation through cavity-assisted interactions. *Phys. Rev. Lett.* **92**, 127902 (2004).
- Childress, L., Taylor, J. M., Sørensen, A. S. & Lukin, M. D. Fault-tolerant quantum repeaters with minimal physical resources and implementations based on single-photon emitters. *Phys. Rev. A* **72**, 052330 (2005).
- Waks, E. & Vučković, J. Dipole induced transparency in drop-filter cavity-waveguide systems. *Phys. Rev. Lett.* **96**, 153601 (2006).
- Ladd, T. D., van Loock, P. K., Nemoto, K., Munro, W. J. & Yamamoto, Y. Hybrid quantum repeater based on dispersive CQED interactions between matter qubits and bright coherent light. *N. J. Phys.* **8**, 184 (2006).
- Birnbaum, K. M. et al. Photon blockade in an optical cavity with one trapped atom. *Nature* **436**, 87–90 (2005).
- Rauschenbeutel, A. et al. Coherent operation of a tunable quantum phase gate in cavity QED. *Phys. Rev. Lett.* **83**, 5166–5169 (1999).
- Nogues, G. et al. Seeing a single photon without destroying it. *Nature* **400**, 239–242 (1999).
- Schuster, D. I. et al. Resolving photon number states in a superconducting circuit. *Nature* **445**, 515–518 (2007).
- Kimble, H. J. in *Cavity Quantum Electrodynamics* (ed. Berman, P.) 213–219 (Academic, San Diego, 1994).
- Akahane, Y., Asano, T., Song, B.-S. & Noda, S. High-Q photonic nanocavity in a two-dimensional photonic crystal. *Nature* **425**, 944–947 (2003).
- Faraon, A. et al. Local quantum dot tuning on photonic crystal chips. *Appl. Phys. Lett.* **90**, 213110 (2007).

17. Gerardot, B. D. *et al.* Contrast in transmission spectroscopy of a single quantum dot. *Appl. Phys. Lett.* **90**, 221106 (2007).
18. Tan, S. M. A computational toolbox for quantum and atomic physics. *J. Opt. B* **1**, 424–432 (1999).
19. Hood, C. J., Chapman, M. S., Lynn, T. W. & Kimble, H. J. Real-time cavity QED with single atoms. *Phys. Rev. Lett.* **80**, 4157–4160 (1998).
20. Auffeves-Garnier, A., Simon, C., Gerard, J. M. & Poizat, J.-P. Giant optical nonlinearity induced by a single two-level system interacting with a cavity in the Purcell regime. *Phys. Rev. A* **75**, 053823 (2007).
21. Englund, D., Faraon, A., Zhang, B., Yamamoto, Y. & Vučković, J. Generation and transfer of single photons on a photonic crystal chip. *Opt. Express* **15**, 5550–5558 (2007).
22. Faraon, A., Waks, E., Englund, D., Fushman, I. & Vučković, J. Efficient photonic crystal cavity-waveguide couplers. *Appl. Phys. Lett.* **90**, 073102 (2007).
23. Reiner, J. E., Smith, W. P., Orozco, L. A., Carmichael, H. J. & Rice, P. R. Time evolution and squeezing of the field amplitude in cavity QED. *J. Opt. Soc. Am. B* **18**, 1911–1921 (2001).
24. Imoto, N., Haus, H. A. & Yamamoto, Y. Quantum nondemolition measurement of the photon number via the optical Kerr effect. *Phys. Rev. A* **32**, 2287–2292 (1985).
25. Vučković, J., Englund, D., Fattal, D., Waks, E. & Yamamoto, Y. Generation and manipulation of nonclassical light using photonic crystals. *Physica E* **32**, 466–470 (2006).

Acknowledgements Financial support was provided by the ONR Young Investigator Award, the MURI Center for photonic quantum information systems (ARO/DTO Program), the Okawa Foundation Faculty Research Grant, and the CIS Seed fund. D.E. and I.F. were also supported by the NDSEG fellowship. Work was performed in part at the Stanford Nanofabrication Facility of NNIN supported by the National Science Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.V. (jela@stanford.edu).

LETTERS

Linear and nonlinear optical spectroscopy of a strongly coupled microdisk–quantum dot system

Kartik Srinivasan^{1†} & Oskar Painter²

Cavity quantum electrodynamics¹, the study of coherent quantum interactions between the electromagnetic field and matter inside a resonator, has received attention as both a test bed for ideas in quantum mechanics and a building block for applications in the field of quantum information processing². The canonical experimental system studied in the optical domain is a single alkali atom coupled to a high-finesse Fabry–Perot cavity. Progress made in this system^{1–5} has recently been complemented by research involving trapped ions⁶, chip-based microtoroid cavities⁷, integrated microcavity-atom-chips⁸, nanocrystalline quantum dots coupled to microsphere cavities⁹, and semiconductor quantum dots embedded in micropillars, photonic crystals and microdisks^{10–12}. The last system has been of particular interest owing to its relative simplicity and scalability. Here we use a fibre taper waveguide to perform direct optical spectroscopy of a system consisting of a quantum dot embedded in a microdisk. In contrast to earlier work with semiconductor systems, which has focused on photoluminescence measurements^{10–14}, we excite the system through the photonic (light) channel rather than the excitonic (matter) channel. Strong coupling, the regime of coherent quantum interactions, is demonstrated through observation of vacuum Rabi splitting in the transmitted and reflected signals from the cavity. The fibre coupling method also allows us to examine the system's steady-state nonlinear properties, where we see a saturation of the cavity–quantum dot response for less than one intracavity photon. The excitation of the cavity–quantum dot system through a fibre optic waveguide is central to applications such as high-efficiency single photon sources^{15,16}, and to more fundamental studies of the quantum character of the system¹⁷.

In the most simplified picture, cavity quantum electrodynamics (cavity QED) consists of a single two-level atom (or equivalent) coupled to an electromagnetic mode of a cavity. A more realistic picture includes dissipative processes, such as cavity loss and atomic decoherence, and excitation of the system, either through the atomic or photonic channel. The observed system response is dependent on both which channel is excited, and what signal is measured. Previous demonstrations of strong coupling between semiconductor microcavities and quantum dots^{10–13} used non-resonant optical pumping to excite the quantum dot stochastically and photoluminescence to probe the system behaviour. In this work, we excite the system coherently through the photonic channel, and detect signatures of cavity–quantum dot coupling in the resonant optical response. Such optical spectroscopy is commonplace in atom Fabry–Perot systems¹, but is more problematic in semiconductor microcavities owing to the comparative difficulty in effectively coupling light into and out of sub-micrometre structures. To effectively interface with the cavity, we use an optical fibre taper waveguide¹⁸. Fibre tapers are standard glass optical fibres that have been heated and stretched to a diameter that

is at or below the wavelength of light, at which point the evanescent field of the guided mode extends into the surrounding air and allows the taper to function as a near-field optic^{7,19–21}.

The experimental set-up used is shown schematically in Fig. 1a, b. At its core is a customized liquid-helium cryostat²² in which piezo-actuated stages have been integrated to enable optical fibre taper

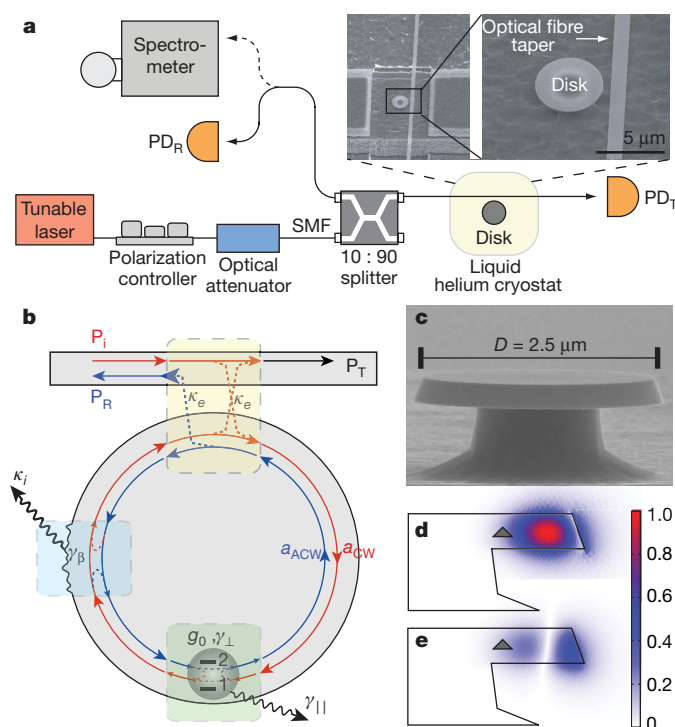


Figure 1 | Experimental apparatus and the cavity–quantum dot system. **a**, Diagram of the experimental set-up, showing a scanning electron microscope (SEM) image of a taper-coupled microdisk. $PD_{R/T}$ are photodetectors for the reflected/transmitted signals and SMF corresponds to single-mode fibre optical links. **b**, Illustration of the coupled microdisk–quantum dot system. Here a_{CW}/a_{CW}^* are the amplitudes for the clockwise/anticlockwise modes, $P_{i/R/T}$ are the incident/reflected/transmitted signals, and κ_e and κ_i correspond respectively to the fibre-to-cavity coupling and the intrinsic cavity field decay rates. For a standing wave WGM, the loaded cavity field decay rate is $\kappa_T = \kappa_i + 2\kappa_e$. **c**, SEM image of one of the small microdisk cavities under study. **d, e**, Finite-element-method simulations of the radial (**d**; E_ρ) and azimuthal (**e**; E_ϕ) electric field components of the $TE_{p=1, m=13}$ mode in cross-section; p denotes the radial order and m the azimuthal mode number. The fields are normalized to the peak radial electric field strength in the cavity. The shaded triangle indicates the estimated quantum dot position in this work.

¹Center for the Physics of Information, ²Thomas J. Watson Sr Laboratory of Applied Physics, California Institute of Technology, Pasadena, California 91125, USA. [†]Present address: Center for Nanoscale Science and Technology, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA.

testing while maintaining a sample temperature as low as 8 K. External cavity tunable lasers optically pump the quantum dot and probe the cavity–quantum dot system near resonance, and fused-fibre couplers direct the cavity’s reflected and transmitted signals to photodetectors and a spectrometer. The overall transmission of the fibre taper link is 50% in this work, and in many cases can be >90%, providing a very-low-loss optical channel to probe the system. This allows for the accurate estimation of quantities such as the average intra-cavity photon number through measurement of the resonant transmission of the taper waveguide when coupled to the cavity.

The system under investigation consists of InAs quantum dots embedded in a GaAs microdisk cavity. The InAs quantum dots are grown in a self-assembled manner with a density of $300\text{--}500\text{ }\mu\text{m}^{-2}$ on top of an InGaAs quantum well (a so-called dot-in-a-well, or DWELL²³). The DWELL structure resides in the middle of a 256-nm-thick GaAs layer that forms the thin planar layer of the microdisk (Fig. 1c). Previous studies of this material²⁴ indicate that isolated emission from single quantum dots at cryogenic temperature can be seen in the wavelength range $\lambda = 1,290\text{--}1,310\text{ nm}$, approximately 50 nm red-shifted from the peak of the quantum dot ensemble emission. Microdisks of diameter $D = 2.5\text{ }\mu\text{m}$ are created through electron beam lithography, plasma dry etching, and wet undercut etching²¹. Finite-element-method simulations (Fig. 1d, e) of the microdisks show that the $\text{TE}_{1,13}$ whispering gallery mode (WGM) is resonant at $\lambda_0 \approx 1,300\text{ nm}$. This optical mode has a radiation-limited quality factor of $Q_{\text{rad}} > 10^8$ in the absence of material absorption and fabrication-induced roughness in the disk shape, and an effective standing wave mode volume $V_{\text{sw}} = 3.2(\lambda_0/n)^3$ for a microdisk refractive index $n \approx 3.4$. The peak coherent coupling rate g_0 between cavity mode and quantum dot for an excitonic state of the type studied here (spontaneous emission lifetime $\tau_{\text{sp}} = 1\text{ ns}$), with optimal placement and dipole orientation, is $g_0/2\pi = 15\text{ GHz}$. Because our quantum dots are not deterministically positioned in the cavity, in contrast to recent studies²⁵, the actual exhibited coupling rate g may be significantly smaller (see Methods). The magnitude of g relative to the system decay rates, κ_T (cavity field decay) and γ_{\perp} (quantum dot dephasing), determines whether the system lies in the perturbative (weak coupling: $g < (\kappa_T, \gamma_{\perp})$) or non-perturbative (strong coupling: $g > (\kappa_T, \gamma_{\perp})$) regime of cavity QED¹.

The process by which we identify a suitable device for studying cavity–quantum dot coupling is described in the Methods section. The main panel of Fig. 2a shows the fibre-taper-collected photoluminescence spectrum from one such device that has been cooled down to 15 K. Optical pumping of the quantum dot is provided by exciting (also through the taper) a blue-detuned higher-order WGM of the disk at $\lambda_p = 982.2\text{ nm}$. The cavity mode, which is fed by background emission processes¹³, is the tall peak at the blue end of the spectrum. The three emission peaks on the red side of the cavity mode are the fine-structure-split²⁶ neutral single exciton lines, X_a and X_b , and the negatively charged single exciton line, X^- .

Further insight into the coupled cavity–quantum dot system from photoluminescence is masked by the limited resolution of our spectrometer (35 pm). In this case, the interesting behaviour of the cavity–quantum dot coupling can be studied by resonant spectroscopy of the cavity mode using a fibre-coupled, narrowband (linewidth <5 MHz) tunable laser. The inset to Fig. 2a shows the taper’s transmission spectrum when it is placed in contact with the side of the microdisk cavity and the cavity modes are detuned from the exciton lines. As has been described in previous work²¹, imperfections on the surface of the microdisk couple the initially degenerate travelling-wave WGMs. If the surface-scattering rate γ_{β} exceeds the total cavity loss rate κ_T , this mode-coupling results in the formation of standing wave modes that are split in frequency. The transmission scan of Fig. 2a inset illustrates this effect in our system, with $\text{TE}_{1,13}$ modes appearing as a resonance doublet with splitting $2\Delta\lambda_{\beta} =$

31 pm. Each mode has a linewidth of $\delta\lambda = 13\text{ pm}$, corresponding to $Q = 10^5$ and $\kappa_T/2\pi = 1.2\text{ GHz}$.

To tune the cavity into resonance with the X_a and X_b exciton lines of the quantum dot, we introduce nitrogen (N_2) gas into the cryostat^{22,27}. As described in ref. 22 and in Methods, this allows for continuous and repeated tuning of the cavity modes over a 4 nm wavelength range. For the first set of measurements, we operate with an input power of 470 pW so that the system remains in a weak driving regime with the estimated bare-cavity intracavity photon number $n_{\text{cav}} = 0.03$. The normalized transmission and reflection spectra over a cavity tuning range of 240 pm are displayed as an intensity image in Fig. 2b, c. Initially, we see a simple shift in the centre wavelength of the cavity doublet mode, but once the cavity mode frequency nears the transition frequency of the higher-energy exciton line (X_a) of the quantum dot, the spectra change dramatically. We see that coupling between the X_a -line and the cavity modes results in a significant spectral splitting (vacuum Rabi splitting) that is evidenced in the characteristic anticrossing within both the transmitted and reflected signals. This anticrossing is indicative of the cavity taking on the character of the quantum dot exciton, and vice versa, when the system becomes strongly coupled. As the cavity is detuned red of the X_a -line, the spectra regain their initial bare-cavity doublet shape. Further tuning brings the cavity modes into resonance with the X_b -line (not shown). Only a small dispersive shift of the cavity modes’ frequencies and no anticrossing is observed in this case, indicating that the X_b state is only weakly coupled to the cavity.

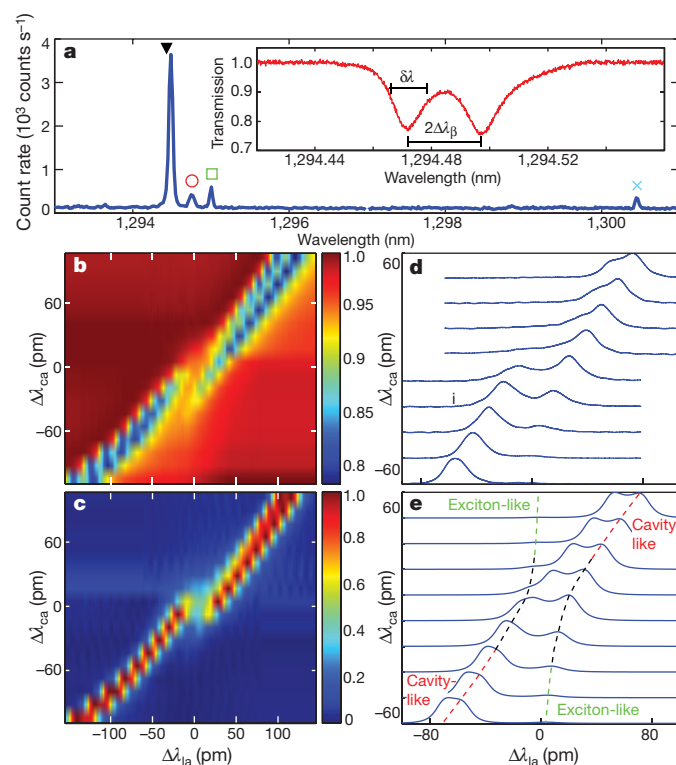


Figure 2 | Reflection and transmission spectra from a strongly coupled microdisk–quantum dot system. **a**, Fibre collected photoluminescence spectrum at a pump power of 30 nW showing the cavity mode (black triangle), X_a (red circle), X_b (green square), and X^- (blue cross) lines. The inset shows a transmission scan of the bare-cavity mode. **b**, **c**, Measured transmission (**b**) and reflection (**c**) spectra as a function of laser–quantum dot detuning ($\Delta\lambda_{\text{la}}$) and cavity–quantum dot detuning ($\Delta\lambda_{\text{ca}}$), where the cavity wavelength is tuned by the N_2 adsorption. Transmission and reflection spectra are normalized to unity. **d**, **e**, Experimental data (**d**) and model plots (**e**) for a series of reflected spectra in the central 120 pm region of cavity tuning. The dashed lines in **e** are guides to the eye for the exciton-like and cavity-like tuning.

Table 1 | Parameters of quantum master equation model

Parameter	V_{tw}	η	$\kappa_{\text{e}}/2\pi$	$\kappa_{\text{i}}/2\pi$	$\gamma_{\text{b}}/2\pi$	ξ	τ_{rad}	$g_{\text{sw1}}/2\pi$	$g_{\text{sw2}}/2\pi$	$\gamma_{\perp}/2\pi$	$\gamma_{\parallel}/2\pi$
	$((\lambda_0/n)^3)$		(GHz)	(GHz)	(GHz)	(rad)	(ns)	(GHz)	(GHz)	(GHz)	(GHz)
Value	6.4	0.21	0.171	0.91	1.99	0.25π	1	2.93	1.21	1.17	0.55

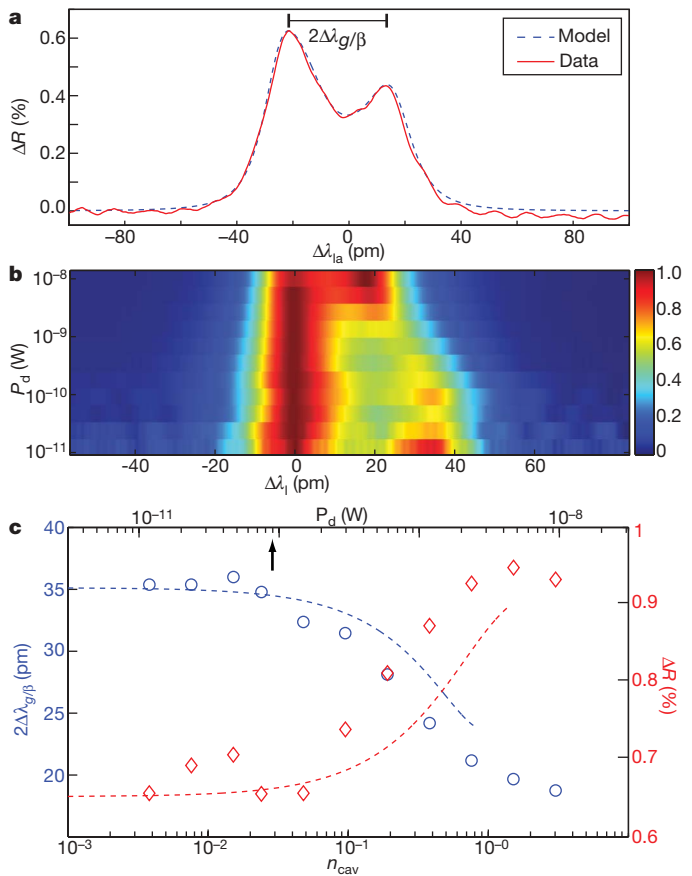
Figure 2d shows a series of reflection scans for a zoomed-in region of cavity tuning, near where the X_{a} -line and the cavity are in resonance. In general, the character of these spectra is complicated by the bimodal nature of WGM cavities. To adequately model the system, we use a quantum master equation²⁸. The model is used to solve for the steady-state reflected and transmitted signals from the cavity as a function of parameters such as cavity–exciton coupling and excitonic dephasing (the bare-cavity properties are known from detuned cavity spectra). One other important parameter is the relative phase, ξ , between the surface scattering and exciton mode coupling. The quantum dot–cavity coupling strength with the standing wave modes, $g_{\text{sw1,2}}$ is modified relative to that for travelling wave WGMs by a factor of $(1 \pm e^{i\xi})/\sqrt{2}$.

A series of reflected spectra produced by the model is shown in Fig. 2e for a set of parameters, listed in Table 1, which best estimates the measured reflected signal intensity, exciton linewidth, relative coupling to the two standing wave modes, and anticrossed splitting. These parameters place the X_{a} exciton state and the $\text{TE}_{1,13}$ WGM in the good-cavity limit ($g > \gamma_{\perp} > \kappa_{\text{T}}$) of the strong coupling regime. We note that the achieved g_{sw1} is about five times smaller than the maximum possible value based on the cavity mode volume, and is probably due to the quantum dot position being suboptimal. We estimate that the quantum dot is located 300–400 nm inwards from the position of peak field strength of the $\text{TE}_{1,13}$ mode (Fig. 1d), with the dipole moment of the X_{a} -line oriented radially and that of the X_{b} -line oriented azimuthally. This picture is consistent with the orthogonal X_{a} – X_{b} polarizations²⁶ and their relative measured coupling strengths.

The rate at which a single exciton can scatter incoming cavity photons is limited, resulting in a saturation in the strongly coupled quantum dot–cavity response for large enough input power. Two parameters used to characterize nonlinear processes in cavity QED are the critical atom number N_0 and the saturation photon number m_0 , which respectively gauge the number of atoms needed to alter the cavity response and the number of photons needed to saturate the atomic transition¹. These parameters are given by $N_0 = 2\kappa_{\text{T}}\gamma_{\perp}/g^2$ and $m_0 = \gamma_{\parallel}\gamma_{\perp}/4g^2$, where γ_{\parallel} and γ_{\perp} are the exciton energy decay rate and total dipole dephasing rate, respectively. In our system, $N_0 = 0.44$ and $m_0 = 0.02$ for the standing wave mode (sw1) that couples most strongly to the quantum dot. This indicates that a single quantum dot strongly affects the cavity response (which Fig. 2 clearly indicates), while even an average intracavity photon number that is less than one can saturate the quantum dot response.

The measured power dependence of the quantum dot–cavity system is shown in Fig. 3, where the cavity is tuned into resonance with the X_{a} -line near the centre of the anticrossing region (scan marked ‘i’ in Fig. 2d), at which point the resonance peaks are nearly equal mixtures of exciton and cavity mode. Figure 3a shows a plot of the measured reflected signal normalized to input power (ΔR) along with the modelled steady-state response of the cavity under weak driving conditions ($n_{\text{cav}} = 0.03$). As the input power to the cavity increases, Fig. 3b shows that the spectral splitting due to cavity–quantum dot interaction ($2\Delta\lambda_{\text{g}}$) begins to diminish as the exciton saturates, and finally reaches a regime where the splitting is nearly half as large and due to surface scattering ($2\Delta\lambda_{\text{b}}$). Figure 3c plots the resulting mode splitting ($2\Delta\lambda_{\text{g/b}}$) and peak ΔR as a function of the optical drive power. Both the measured splitting and reflected signal begin to saturate towards their bare-cavity values for $n_{\text{cav}} \gtrsim 0.1$. The model based on the quantum master equation (dotted lines) predicts very similar behaviour, with a smooth, extended region of saturation, as expected for a single dipole with quantum fluctuations^{3,29}.

Use of an optical-fibre-based waveguide to efficiently probe the microcavity–quantum dot system opens up many interesting possibilities for future devices and studies. In particular, excitation and collection through the optical channel allows for high resolution spectral and temporal studies of individual quantum dot dynamics and a direct probe of the intra-cavity field. Studies of the quantum fluctuations of the strongly coupled system¹⁷, through field and intensity correlations of the optical signal, are also now possible. An immediate application is the creation of an efficient fibre-coupled single-photon source, whereas from a long-term perspective, the fibre interface could perhaps serve as a means to transfer quantum information to and from the quantum dot. In comparison, atomic systems have the considerable advantages of homogeneity, much lower dephasing, and an energy level structure compatible with more complex manipulations of the quantum system. Nitrogen-vacancy centres in diamond^{9,30} have been viewed as a system that can provide

**Figure 3 | Power dependence of the quantum dot–microcavity system.**

a, Reflection spectrum from the quantum dot–microdisk system near resonance (position ‘i’ in Fig. 2) under weak driving. The solid red line is the measured reflected power normalized to input power; the dashed blue line is a quantum master equation model of the system. **b**, Normalized (to unity) reflected signal of **a** as a function of drive strength (dropped power in the bare cavity, P_{d}) and detuning from the short-wavelength resonance peak ($\Delta\lambda_{\text{l}}$). **c**, Measured and modelled saturation of the mode splitting and peak reflected signal level versus drive strength (n_{cav} (bottom axis); P_{d} (top axis)). The model is only plotted up to a drive power of $n_{\text{cav}} = 1$ owing to size limitations on the cavity mode Fock space that can be simulated. Bold arrow indicates the drive strength used for **a** and Fig. 2b–e.

some of the beneficial aspects of cold atoms. The measurement apparatus described here is equally applicable to this and other systems, and we are hopeful that it can be built on to further progress the development of solid-state cavity QED nodes with microchip scalability.

METHODS SUMMARY

Device identification consists of the following procedure. An array of 50 microdisks is fabricated per sample, with disk diameter nominally equal to $D = 2.5 \mu\text{m}$. Nanometre-scale fluctuations in disk diameter cause the $\text{TE}_{1,13}$ mode wavelength to vary over a 1,290–1,310 nm range from device to device in the array. Each microdisk is optically pumped through the fibre taper and on resonance with one of its WGMs in the 980 nm wavelength band²⁴. This selectively excites quantum dots that lie in the disk periphery and overlap with the $\text{TE}_{1,13}$ mode. For those devices in which isolated quantum dot emission is observed, we examine the spectral position of the $\text{TE}_{1,13}$ mode relative to the quantum dot states through photoluminescence and cavity transmission. A digital wet etching process²⁵ provides a cavity mode blue shift of 0.8 nm per cycle. This wet etch is repeated until the cavity mode, of a chosen device on the sample, lies blue (and within 1 nm) of the desired quantum dot exciton lines. N_2 adsorption is then used to red-shift the mode into resonance with a given exciton line.

The numerically calculated travelling wave mode volume of the $\text{TE}_{1,13}$ WGM is $V_{\text{tw}} = 6.4(\lambda_0/n)^3$ for the $D = 2.5 \mu\text{m}$ microdisks studied in this work. The coherent coupling rate of the exciton to the travelling wave mode is given by $g_{\text{tw}} = \eta(3c\lambda_0^2/8\pi n^2 \epsilon_{\text{sp}} V_{\text{tw}})^{1/2}$, where η accounts for the position and orientation of the exciton dipole ($\eta = 1$ for an exciton dipole oriented parallel with, and positioned at, the peak of the cavity mode electric field).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 July; accepted 11 September 2007.

- Kimble, H. J. Strong interactions of single atoms and photons in cavity QED. *Phys. Scripta T* **76**, 127–137 (1998).
- Mabuchi, H. & Doherty, A. C. Cavity quantum electrodynamics: coherence in context. *Science* **298**, 1372–1377 (2002).
- Hood, C. J., Chapman, M. S., Lynn, T. W. & Kimble, H. J. Real-time cavity QED with single atoms. *Phys. Rev. Lett.* **80**, 4157–4160 (1998).
- Hennrich, M., Legero, T., Kuhn, A. & Rempe, G. Vacuum-stimulated Raman scattering based on adiabatic passage in a high-finesse optical cavity. *Phys. Rev. Lett.* **85**, 4872–4875 (2000).
- Boca, A. *et al.* Observation of the vacuum Rabi spectrum for one trapped atom. *Phys. Rev. Lett.* **93**, 233603 (2004).
- Keller, M., Lange, B., Hayaska, K., Lange, W. & Walther, H. Continuous generation of single photons with controlled waveform in an ion-trap cavity system. *Nature* **431**, 1075–1078 (2004).
- Aoki, T. *et al.* Observation of strong coupling between one atom and a monolithic microresonator. *Nature* **443**, 671–674 (2006).
- Colombe, Y. *et al.* Strong atom-field coupling for Bose-Einstein condensates in an optical cavity on a chip. Preprint at (<http://arxiv.org/abs/0706.1390>) (2007).
- Park, Y.-S., Cook, A. K. & Wang, H. Cavity QED with diamond nanocrystals and silica microspheres. *Nano Lett.* **6**, 2075–2079 (2006).
- Reithmaier, J. P. *et al.* Strong coupling in a single quantum dot-semiconductor microcavity system. *Nature* **432**, 197–200 (2004).
- Yoshie, T. *et al.* Vacuum Rabi splitting with a single quantum dot in a photonic crystal nanocavity. *Nature* **432**, 200–203 (2004).
- Peter, E. *et al.* Exciton photon strong-coupling regime for a single quantum dot embedded in a microcavity. *Phys. Rev. Lett.* **95**, 067401 (2005).
- Hennessy, K. *et al.* Quantum nature of a strongly coupled single quantum dot-cavity system. *Nature* **445**, 896–899 (2007).
- Khitrova, G., Gibbs, H. M., Kira, M., Koch, S. W. & Scherer, A. Vacuum Rabi splitting in semiconductors. *Nature Phys.* **2**, 81–90 (2006).
- Michler, P. *et al.* A quantum dot single-photon turnstile device. *Science* **290**, 2282–2285 (2000).
- Santori, C., Fattal, D., Vuckovic, J., Solomon, G. & Yamamoto, Y. Indistinguishable photons from a single-photon device. *Nature* **419**, 594–597 (2002).
- Birnbaum, K. M. *et al.* Photon blockade in an optical cavity with one trapped atom. *Nature* **436**, 87–90 (2005).
- Knight, J. C., Cheung, G., Jacques, F. & Birks, T. A. Phase-matched excitation of whispering-gallery-mode resonances by a fiber taper. *Opt. Lett.* **22**, 1129–1131 (1997).
- Spillane, S. M., Kippenberg, T. J., Painter, O. J. & Vahala, K. J. Ideality in a fiber-taper-coupled microresonator system for application to cavity quantum electrodynamics. *Phys. Rev. Lett.* **91**, 043902 (2003).
- Srinivasan, K., Barclay, P. E., Borselli, M. & Painter, O. Optical-fiber-based measurement of an ultrasmall volume, high-Q photonic crystal microcavity. *Phys. Rev. B* **70**, 081306R (2004).
- Srinivasan, K. *et al.* Optical loss and lasing characteristics of high-quality-factor AlGaAs microdisk resonators with embedded quantum dots. *Appl. Phys. Lett.* **86**, 151106 (2005).
- Srinivasan, K. & Painter, O. Optical fiber taper coupling and high-resolution wavelength tuning of microdisk resonators at cryogenic temperatures. *Appl. Phys. Lett.* **90**, 031114 (2007).
- Liu, G. T. *et al.* The influence of quantum-well composition on the performance of quantum dot lasers using InAs/InGaAs dots-in-a-well (DWELL) structures. *IEEE J. Quant. Electron.* **36**, 1272–1279 (2000).
- Srinivasan, K., Painter, O., Stintz, A. & Krishna, S. Single quantum dot spectroscopy using a fiber taper waveguide near-field optic. *Appl. Phys. Lett.* **91**, 091102 (2007).
- Badolato, A. *et al.* Deterministic coupling of single quantum dots to single nanocavity modes. *Science* **308**, 1158–1161 (2005).
- Kulakovski, V. D. *et al.* Fine structure of biexciton emission in symmetric and asymmetric CdSe/ZnSe single quantum dots. *Phys. Rev. Lett.* **82**, 1780–1783 (1999).
- Mosor, S. *et al.* Scanning a photonic crystal slab nanocavity by condensation of xenon. *Appl. Phys. Lett.* **87**, 141105 (2005).
- Srinivasan, K. & Painter, O. Mode coupling and cavity-quantum-dot interactions in a fiber-coupled microdisk cavity. *Phys. Rev. A* **75**, 023814 (2007).
- Savage, C. M. & Carmichael, H. J. Single-atom optical bistability. *IEEE J. Quant. Electron.* **24**, 1495–1498 (1988).
- Santori, C. *et al.* Coherent population trapping of single spins in diamond under optical excitation. *Phys. Rev. Lett.* **97**, 247401 (2006).

Acknowledgements We thank S. Krishna and A. Stintz for providing quantum dot material growth. This work was supported by the Charles L. Powell Foundation and the Center for the Physics of Information at Caltech.

Author Contributions Both K.S. and O.P. contributed to all aspects of this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to O.P. (opainter@caltech.edu).

METHODS

Cavity tuning. Nitrogen is released into the chamber in discrete 5 s increments, with the flow rate adjusted so that a tuning level of ~ 10 pm per step is achieved. At temperatures above 28 K, the N_2 can be removed from the disk surface and the cavity mode reset back to its original wavelength, allowing for repeated tuning cycles.

Transmission/reflection measurements. A narrowband single-mode laser (linewidth < 5 MHz) with continuous wavelength tuning in the $\lambda = 1,300$ nm band is used to probe the cavity–quantum dot system. The transmitted and reflected laser signals are detected by thermal-electric-cooled (1 kHz bandwidth) and liquid-nitrogen-cooled (150 Hz bandwidth) InGaAs photodetectors, respectively. The photodetected signals are low-pass filtered (30 Hz cut-off) and the wavelength scans are averaged 10–20 times to produce the spectra of Fig. 2. Photoluminescence is dispersed through a 550 mm Czerny–Turner spectrometer and detected on a 512 element liquid-nitrogen-cooled InGaAs array ($25\ \mu\text{m} \times 500\ \mu\text{m}$ pixel size yields a resolution of 35 pm).

Quantum master equation simulations. Reference 28 presents an appropriate model for our system. We numerically solve the steady-state quantum master equation for the system's density matrix, from which the transmitted and reflected spectra from the cavity are generated. A Fock space dimension of 6 for each cavity mode was used in modelling the drive power dependence of the system shown in Fig. 3. The expectation of the commutation between creation and annihilation operators for each mode was calculated to ensure accuracy of the simulation.

LETTERS

A crystallizing dense magma ocean at the base of the Earth's mantle

S. Labrosse¹, J. W. Hernlund^{2,†} & N. Coltice^{1,3}

The distribution of geochemical species in the Earth's interior is largely controlled by fractional melting and crystallization processes that are intimately linked to the thermal state and evolution of the mantle. The existence of patches of dense partial melt at the base of the Earth's mantle¹, together with estimates of melting temperatures for deep mantle phases² and the amount of cooling of the underlying core required to maintain a geodynamo throughout much of the Earth's history³, suggest that more extensive deep melting occurred in the past. Here we show that a stable layer of dense melt formed at the base of the mantle early in the Earth's history would have undergone slow fractional crystallization, and would be an ideal candidate for an unsampled geochemical reservoir hosting a variety of incompatible species (most notably the missing budget of heat-producing elements) for an initial basal magma ocean thickness of about 1,000 km. Differences in ¹⁴²Nd/¹⁴⁴Nd ratios between chondrites and terrestrial rocks⁴ can be explained by fractional crystallization with a decay timescale of the order of 1 Gyr. These combined constraints yield thermal evolution models in which radiogenic heat production and latent heat exchange prevent early cooling of the core and possibly delay the onset of the geodynamo to 3.4–4 Gyr ago⁵.

The survival of a layer of melt formed at the base of the Earth's mantle early in its history (Fig. 1a) will have depended on whether it was both gravitationally and chemically stable. Gravitational stability

is satisfied if the melt formed in Earth's deep mantle is more dense than mantle solids on account of a modest enrichment in iron relative to magnesium and a small⁶ or negative⁷ molar volume change for silicate melting at high pressures. The low viscosity of such a melt layer ensures vigorous convection and mixing that maintains nearly isentropic conditions (Supplementary Information) and provides a large effective volume for chemical interaction with the core. Thus the chemical stability of a basal melt layer largely hinges on the capacity of the core to come to equilibrium without entirely consuming the layer or removing those chemical components that allow the layer to remain gravitationally stable.

Assuming stability of such a primordial basal melt layer, a simple model for its evolution coupled to the core and the overlying solid mantle can be constructed by assuming an isentropic temperature in the melt layer, an isentropic core³, and a thermal boundary layer at the base of the solid mantle in which the temperature varies linearly with depth (Fig. 2a):

$$4\pi a^2 k \frac{T_L - T_M}{\delta} = -(M_m C_{pm} + M_c C_{pc}) \frac{dT_L}{dt} + H(t) - 4\pi a^2 \rho \Delta S L \frac{da}{dt} \quad (1)$$

with a the upper radius of the melt layer, T_L its liquidus temperature, T_M the temperature above the solid mantle boundary layer, δ the thickness of the boundary layer (assumed constant; Supplementary Information), M_m and M_c the respective masses of the melt layer and

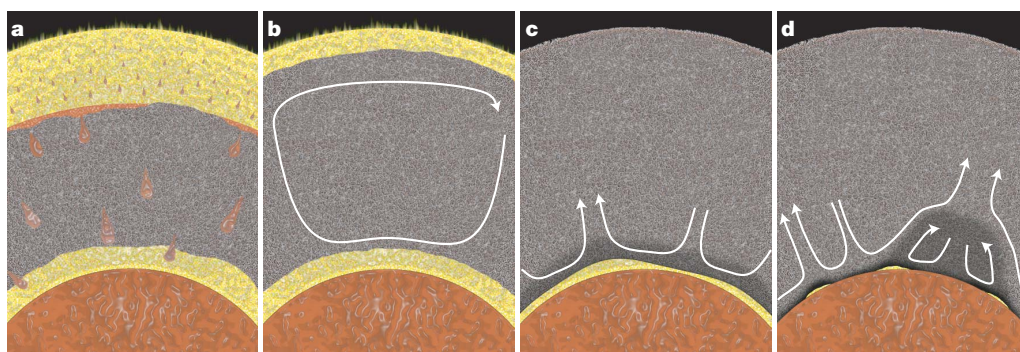


Figure 1 | Schematic illustration of the formation and evolution of a dense basal magma ocean. **a**, Iron-rich liquid descends as a rain of droplets in the shallower magma ocean, accumulates on top of the solid mantle and undergoes diapiric instability and rapid transport to the core. **b**, The molten layers formed at the top and bottom of the mantle crystallize, and deposit material onto a solid mantle layer that grows upward at the top and downward at the bottom at two vastly different rates. **c**, After the surface magma ocean has fully crystallized, the slowly cooling basal melt layer fractionally crystallizes increasingly Fe-enriched solids that are deposited upwards onto the bottom of the solid mantle. **d**, After a substantial part of

the basal magma ocean has frozen, the solid that forms may itself contain enough dense components to become stable against complete entrainment in the solid mantle, hence forming piles under upwelling currents. The remaining thin mushy layer of melt is thicker where mantle flow converges along the core–mantle boundary, leading to seismically detectable ultralow-velocity zones. Solid-state convection in the mantle (white arrows in **b**, **c**, **d**) controls the rate of crystallization of the bottom magma ocean and the possible entrainment of FeO-enriched dense material accumulating at the base of the solid mantle (dark grey in **c** and **d**).

¹Laboratoire des sciences de la Terre, Ecole Normale Supérieure de Lyon, Université de Lyon, CNRS UMR 5570, 46 Allée d'Italie, 69364 Lyon Cedex 07, France. ²Équipe de Dynamique des Fluides Géologiques, Institut de Physique du Globe de Paris, 4 place Jussieu, 75252 Paris Cedex 05, France. ³Laboratoire des sciences de la Terre, Université Lyon 1, Université de Lyon, CNRS UMR 5570, 2 rue Raphaël Dubois, 69622 Villeurbanne Cedex, France. [†]Present address: Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

the core, C_{pm} and C_{pC} their respective specific heats, H the radiogenic heat production in the layer, ρ the density of the melt, k the thermal conductivity of the solid mantle, and ΔS the specific entropy of melting.

Recent shock measurements suggest that a deep melt layer will maintain a relatively large adiabatic gradient relative to the change in liquidus temperature with depth, which would ensure that crystallization should proceed downward from the top of the layer on cooling⁷. The thermal evolution is coupled to the crystallization of the melt layer via the phase diagram, which is not well constrained for the deep mantle. For purposes of illustration, we use an idealized solid-solution (Fig. 2b), where an MgO-rich end-member (denoted A) has a higher melting temperature T_A than the melting temperature T_B of a dense FeO-rich component (denoted B). On fractional crystallization, the melt is gradually enriched in component B, and denoting by ξ_L and ξ_S the mass fraction of B in the melt and the solid, respectively, conservation of chemical species gives:

$$\frac{d\xi_L}{dt} = -\frac{3a^2\Delta\xi}{a^3-b^3} \frac{da}{dt} \quad (2)$$

with $\Delta\xi = \xi_L - \xi_S$ the enrichment of the melt in the dense component relative to the solid, and b the radius of the core. The linearized phase diagram is used to relate $d\xi_L/dt$ to dT_L/dt , and we obtain two coupled equations for T_L and a that can be solved numerically from any time where both quantities are known.

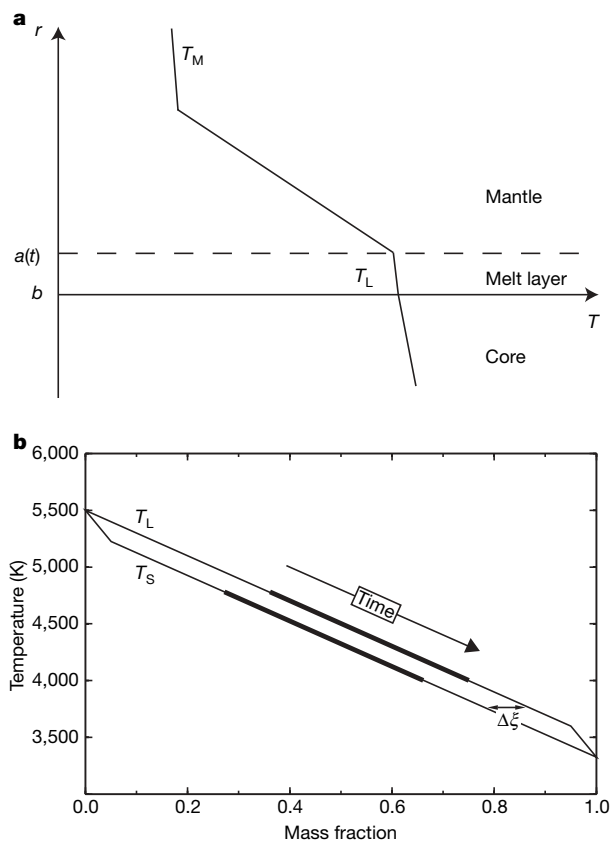


Figure 2 | Sketch of the idealized model. **a**, Model for the decrease of temperature T as function of radius r , with isentropic profiles in the core ($r < b$) and melt layer ($b < r < a(t)$) up to the liquidus temperature of mantle material T_L at the freezing interface, overlaid by a boundary layer in the solid mantle ($r > a(t)$) where the temperature decreases linearly with radius to reach the lower mantle temperature T_M . **b**, Idealized phase diagram for a solid solution, for example $\text{Mg}_x\text{Fe}_{1-x}\text{O}$ or $\text{Mg}_x\text{Fe}_{1-x}\text{SiO}_3$, the liquidus and solidus temperature T_L and T_S having $\Delta\xi$ as difference in mass fraction of the light component. The thickened lines represent the region of the phase diagram that is covered during the evolution presented in Fig. 3.

In this idealized case, the thickness of the melt layer decreases exponentially with time while the temperature decreases linearly (Fig. 3), consistent with an approximate analytic solution for $a - b \ll b$ and a constant cooling timescale of $\tau_C = M_C C_{pC} \Delta\xi (T_A - T_B) / (4\pi b^2 k (T_L - T_M) / \delta - H)$, as suggested by the energy balance (Fig. 3). In this case, τ_C is the e-fold time for the decay of the layer thickness, and $\Delta\xi (T_A - T_B) / \tau_C$ is the linear decay constant for the temperature. It is primarily controlled by the product of the heat capacity of the core and the temperature range ($T_A - T_B$) over which a liquid can exist in the phase diagram. Although the

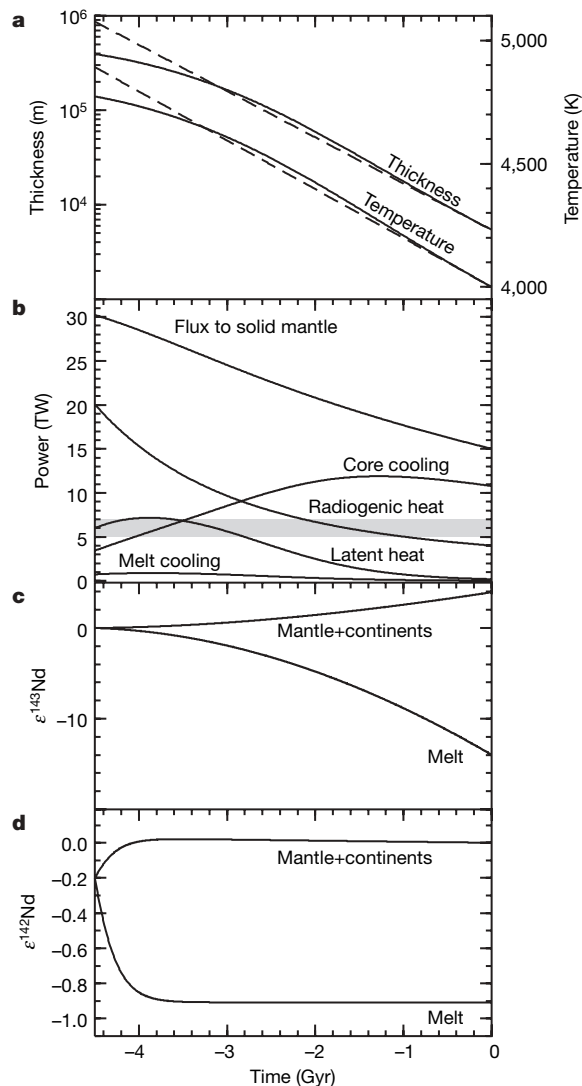


Figure 3 | Evolution of the preferred model as a function of time. **a**, Thickness (left scale), and temperature (right scale) of the melt layer. **b**, Energy budget. **c**, **d**, $\epsilon^{143}\text{Nd}$ (**c**) and $\epsilon^{142}\text{Nd}$ (**d**) of the solid and melt reservoirs predicted by the model of crystallization of the basal magma ocean. The solid lines in **a** and **b** are obtained by numerical integration of equations (1) and (2), whereas the dashed lines are a corresponding approximate analytic solution (Supplementary Information). The light grey area in **b** gives the range of core cooling that must be exceeded in order to support thermal convection in the core, and shows that the onset of the geodynamo is delayed to 4–3.4 Gyr ago in this example. The Sm–Nd calculations are made using $D_{\text{Sm}} = 0.05$, $D_{\text{Nd}} = 0.02$ (ref. 18) for partition coefficients between melt and solid for Sm and Nd, an initial mass of melt $M_{\text{mi}} = 0.78 \times 10^{24}$ kg, a decay timescale for the melt $\tau_C = 887$ Myr, an initial melt fraction $F = 0.8$ and a time of magma ocean formation after formation of calcium aluminium inclusions (CAIs) $t_i = 30$ Myr, corresponding to the thermal evolution shown in **a** and **b**. Other parameters for the thermal evolution model are given in Supplementary Information.

physical parameters appearing in this problem are ill-constrained, an order-of-magnitude estimate yields τ_c of the order of 1 Gyr.

Chemical interaction with the core is neglected in this simple example, although it might be expected to have an important effect on the evolution of the melt layer. For example, an enrichment in iron oxides inside the silicate melt layer with time may favour the formation of FeO and FeSi (ref. 8) liquids when reacted with metallic Fe; these products may or may not be miscible in the core, depending on whether or not it is initially saturated or whether a stratified light-element-enriched layer forms at the top of the core⁹. Removal of such products into the core could accelerate freezing of the basal layer, by consuming a presumed low-melting-temperature Fe-rich component in the melt. On the other hand, cooling of a core whose surface is saturated would favour expulsion of light elements from the core if they become relatively less soluble in metallic Fe as temperature decreases¹⁰; such expulsion could possibly enrich the melt layer in low-melting-temperature components via the reverse reaction and prolong its crystallization. Although the equilibrium processes governing this kind of interaction are not well known, and also depend on a variety of unknown initial conditions that are set by the formation of the core in the early Earth, the plausibility of a slowly crystallizing basal melt layer can additionally be evaluated using other independent constraints.

Fractional crystallization of a large body of liquid in the deep mantle will have important consequences for the distribution of chemical species in Earth's mantle. It is well known that the continental crust and the depleted mantle are not geochemically complementary to one another with respect to the entire mass of Earth's mantle. In particular, if the Earth has a chondritic composition for refractory elements, about 20–30% of the budget of very incompatible elements (and hence heat-producing species such as uranium, thorium and potassium) must be sequestered in a 'hidden reservoir'¹¹, such as a pristine layer in the lower mantle¹², or in a D' layer made of segregated oceanic crust¹³. The mass balance of coupled ⁴⁰K and ⁴⁰Ar requires that the hidden reservoir, rich in ⁴⁰K, also contains 20–60% of the ⁴⁰Ar budget¹⁴. As a consequence, this reservoir must have remained undegassed for several billion years; however, it cannot be pristine as its Nb/Ta and Nb/La ratios must be higher than those of the bulk silicate Earth^{15,16}. Because Nb is more incompatible than Ta and La, the hidden reservoir should be derived from a melt. Additionally, every measured Earth sample has excess ¹⁴²Nd ($\epsilon^{142}\text{Nd} \geq 0$) relative to chondrites ($\epsilon^{142}\text{Nd}$ of -0.2)⁴. As the Earth is expected to be chondritic for refractory elements, the ¹⁴²Nd signature of sampled rocks has to be balanced by a hidden and unsampled reservoir in the deep mantle having an even more negative $\epsilon^{142}\text{Nd}$ and a low Sm/Nd ratio. The hidden reservoir has to be formed within the first 100 Myr of Earth's history at most, because ¹⁴²Nd is produced by decay of ¹⁴⁶Sm with a half-life of 103 Myr (ref. 17).

The basal magma ocean is a unique candidate for the hidden geochemical reservoir. Therefore, we can use the balance of incompatible elements in addition to ¹⁴²Nd/¹⁴⁴Nd and ¹⁴³Nd/¹⁴⁴Nd data to constrain its thermal evolution. We take partition coefficients between (Mg,Fe)SiO₃ perovskite (the assumed liquidus phase) and melt from experimental results¹⁸. Using an exponentially decaying mass for the melt layer as discussed previously, we solve analytically the evolution of the whole Sm–Nd system constrained by the observed present-day ¹⁴²Nd/¹⁴⁴Nd and ¹⁴³Nd/¹⁴⁴Nd ratios of the bulk mantle (Supplementary Information and Fig. 3) starting from a chondritic Earth. Because the melt is denser and much less viscous than the overlying solid mantle, entrainment into the mantle is negligible and therefore the geochemical signal of the melt layer is unsampled¹⁷. Models predicting 20–30% of the total chondritic uranium budget in the deep melt today and realistic present-day values of ¹⁴²Nd/¹⁴⁴Nd and ¹⁴³Nd/¹⁴⁴Nd for the solid mantle require an initial 850 ± 150 km thick basal magma ocean to be extracted by 80 ± 5% melting of the lowermost mantle (Supplementary

Information). The residue of melting ultimately mixes back into the overlying solid mantle (Fig. 1b, c). Nevertheless, fractional crystallization produces crystals that are highly depleted in incompatible elements and returns a Sm/Nd signal that is very dilute with respect to the overall mantle. Predictions can be made for the composition of the solid mantle and the underlying melt layer using the above geochemical constraints. The initial composition of the melt is computed from 80% batch melting of primitive mantle¹⁹ to produce a melt layer 850 km thick. With a fractional crystallization model, we obtain the present-day composition of the melt, which is slightly richer in incompatible elements than the continental crust (Fig. 4) and accounts for about 20% of the total budget. Applying a mass balance with the continental crust²⁰ and the remaining melt, we can then predict a composition of the solid mantle (Fig. 4) that is consistent with estimates of the shallow mantle²¹ and with the trace-element pattern of normal mid-ocean-ridge basalts²².

The crystallizing solids can exhibit increased Fe/Mg with time because of magma differentiation, and eventually their intrinsic density excess relative to the overlying mantle may become large enough to stabilize them against entrainment by mantle convection¹². In this scenario, the shape of the denser crystalline piles is determined by the viscous coupling with the overlying mantle, with downgoing slabs pushing aside the Fe-rich rocks that tend to accumulate under upwelling currents. The existence of such piles has already been inferred from seismological studies²³, in which a counter-circulation developing within them tends to sweep residual melt to their edges²⁴ (Fig. 1d). A small fraction of the dense cumulates can be entrained but this should produce no geochemical signal, as incompatible elements stay within the melt rather than within the cumulates. Noble gases may constitute an exception because of the intense outgassing that occurred at the surface of the overlying solid mantle (more than 99% degassing for non-radiogenic helium, neon and argon²⁵).

Another implication of the basal magma ocean is that the heat flow into the solid mantle has a contribution from heat production in the melt layer, in addition to the cooling of the core (equation (1)).

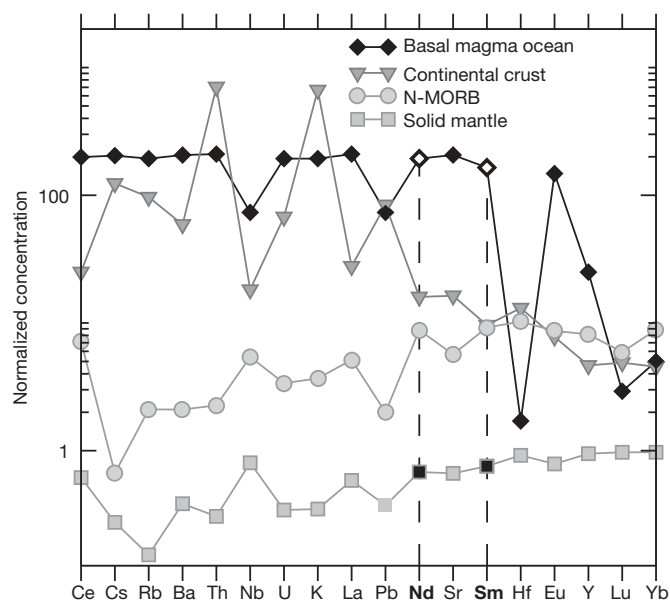


Figure 4 | Predicted element concentrations in the present-day basal magma ocean and solid mantle. Also shown for comparison are patterns for the continental crust²⁰ and the average normal mid-ocean-ridge basalt (N-MORB)²². All concentrations have been normalized by those of the bulk silicate Earth¹⁹. The present concentration in the melt results from fractional crystallization (calculated using the partition coefficients of ref. 18) and a solid mantle composition calculated from mass balance with the continental crust and the basal magma ocean. Nd and Sm are emphasized to highlight their difference in partition coefficients, responsible for the evolution of ¹⁴²Nd and ¹⁴³Nd (Fig. 3).

Hence, estimates of core cooling from seismic observations of a double-crossing of the post-perovskite phase boundary^{26,27} might be revised to lower values by up to 4–6 TW. More importantly, both radiogenic heating and the contribution from latent heat of crystallization were significantly larger in the early Earth, limiting core heat loss to values lower than the minimal value (5–7 TW) required for thermal convection (Fig. 3). As the inner core would only appear around 1 Gyr ago with such a thermal history³, the early geodynamo cannot be supported by compositional convection and can only start at the onset of thermal convection, between 4 and 3.4 Gyr ago. The delayed onset of the geodynamo would allow implantation of some of Earth's early atmosphere in lunar soil⁵, and is consistent with the oldest known palaeomagnetic record²⁸.

Received 21 May; accepted 2 October 2007.

- Williams, Q. & Garnero, E. J. Seismic evidence for partial melt at the base of the Earth's mantle. *Science* **273**, 1528–1530 (1996).
- Zerr, A., Diegeler, A. & Boehler, R. Solidus of Earth's deep mantle. *Science* **281**, 243–246 (1998).
- Labrosse, S. Thermal and magnetic evolution of the Earth's core. *Phys. Earth Planet. Inter* **140**, 127–143 (2003).
- Boyet, M. & Carlson, R. W. ¹⁴²Nd Evidence for early (> 4.53 Ga) global differentiation of the silicate Earth. *Science* **309**, 576–581 (2005).
- Ozima, M. et al. Terrestrial nitrogen and noble gases in lunar soils. *Nature* **436**, 655–659 (2005).
- Stixrude, L. & Karki, B. Structure and freezing of MgSiO₃ liquid in Earth's lower mantle. *Science* **310**, 297–299 (2005).
- Mosenfelder, J. P., Asimow, P. D. & Ahrens, T. J. Thermodynamic properties of Mg₂SiO₄ liquid at ultra-high pressures for shock measurements to 200 GPa on forsterite and wadsleyite. *J. Geophys. Res.* **112**, B06208, doi:10.1029/2006JB004364 (2007).
- Knittle, E. & Jeanloz, R. Earth's core-mantle boundary: Results of experiments at high pressures and temperatures. *Science* **251**, 1438–1443 (1991).
- Braginsky, S. I. Dynamics of the stably stratified ocean at the top of the core. *Phys. Earth Planet. Inter* **111**, 21–34 (1999).
- Buffett, B. A. Earth's core and the geodynamo. *Science* **288**, 2007–2012 (2000).
- Hofmann, A. W. Mantle geochemistry: The message from oceanic volcanism. *Nature* **385**, 219–229 (1997).
- Davaille, A. Simultaneous generation of hotspots and superswells by convection in a heterogeneous planetary mantle. *Nature* **402**, 756–760 (1999).
- Coltice, N. & Ricard, Y. Geochemical observations and one layer mantle convection. *Earth Planet. Sci. Lett.* **174**, 125–137 (1999).
- Allègre, C. J., Hofmann, A. & O'Nions, K. The argon constraints on mantle structure. *Geophys. Res. Lett.* **23**, 3555–3557 (1996).
- Sun, S. S. & McDonough, W. F. in *Magmatism in the Ocean Basins* (eds Saunders, A. & Norry, M.) 313–345 (Spec. Publ. Vol. 42, Geol. Soc. Lond., 1989).
- Rudnick, R. L., Barth, M., Horn, I. & McDonough, W. F. Rutile-bearing refractory eclogites: Missing link between continents and depleted mantle. *Science* **287**, 278–281 (2000).
- Boyet, M. & Carlson, R. W. A new geochemical model for the Earth's mantle inferred from ¹⁴⁶Sm–¹⁴²Nd systematics. *Earth Planet. Sci. Lett.* **250**, 254–268 (2006).
- Corgne, A., Liebske, C., Wood, B. J., Rubie, D. C. & Frost, D. J. Silicate perovskite-melt partitioning of trace elements and geochemical signature of a deep perovskitic reservoir. *Geochim. Cosmochim. Acta* **69**, 485–496 (2005).
- McDonough, W. F. & Sun, S. S. The composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).
- Rudnick, R. L. & Fountain, D. M. Nature and composition of the continental crust: A lower crustal perspective. *Rev. Geophys.* **33**, 267–309 (1995).
- Salter, V. J. M. & Stracke, A. Composition of the depleted mantle. *Geochem. Geophys. Geosyst.* **5**, Q05004, doi:10.1029/2003GC000597 (2004).
- Hofmann, A. W. Chemical differentiation of the Earth: The relationship between mantle, continental crust and oceanic crust. *Earth Planet. Sci. Lett.* **90**, 297–314 (1988).
- Wen, L. X., Silver, P., James, D. & Kuehnel, R. Seismic evidence for a thermo-chemical boundary at the base of the Earth's mantle. *Earth Planet. Sci. Lett.* **189**, 141–153 (2001).
- Thorne, M. S. & Garnero, E. J. Inferences on ultralow-velocity zone structure from a global analysis of SPdKS waves. *J. Geophys. Res.* **109**, B08301, doi:10.1029/2004JB003010 (2004).
- Allègre, C. J., Staudacher, T. & Sarda, P. Rare gas systematics: Formation of the atmosphere, evolution and structure of the Earth's mantle. *Earth Planet. Sci. Lett.* **81**, 127–150 (1987).
- Hernlund, J. W., Thomas, C. & Tackley, P. J. A doubling of the post-perovskite phase boundary and structure of the Earth's lowermost mantle. *Nature* **434**, 882–886 (2005).
- Lay, T., Hernlund, J., Garnero, E. J. & Thorne, M. S. A post-perovskite lens and D'' heat flux beneath the central Pacific. *Science* **314**, 1272–1276 (2006).
- Tarduno, J. A., Cottrell, R. D., Watkeys, M. K. & Bauch, D. Geomagnetic field strength 3.2 billion years ago recorded by single silicate crystals. *Nature* **446**, 657–660 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Stevenson for comments that helped us to considerably sharpen this paper. M. Moreira, C. Jaupart and M. Jellinek also provided valuable feedback. This research was supported by the SEDIT programme of INSU, the French Ministry of Research and a CIAR postdoctoral fellowship.

Author Contributions All authors contributed equally to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.L. (stephane.labrosse@ens-lyon.fr).

LETTERS

Host–parasite ‘Red Queen’ dynamics archived in pond sediment

Ellen Decaestecker^{1,3}, Sabrina Gaba^{4,5}, Joost A. M. Raeymaekers^{1,2}, Robby Stoks¹, Liesbeth Van Kerckhoven¹, Dieter Ebert^{4*} & Luc De Meester^{1*}

Antagonistic interactions between hosts and parasites are a key structuring force in natural populations, driving coevolution^{1,2}. However, direct empirical evidence of long-term host–parasite coevolution, in particular ‘Red Queen’ dynamics—in which antagonistic biotic interactions such as host–parasite interactions can lead to reciprocal evolutionary dynamics—is rare^{3–5}, and current data, although consistent with theories of antagonistic coevolution, do not reveal the temporal dynamics of the process⁶. Dormant stages of both the water flea *Daphnia* and its micro-parasites are conserved in lake sediments, providing an archive of past gene pools. Here we use this fact to reconstruct rapid coevolutionary dynamics in a natural setting and show that the parasite rapidly adapts to its host over a period of only a few years. A coevolutionary model based on negative frequency-dependent selection, and designed to mimic essential aspects of our host–parasite system, corroborated these experimental results. In line with the idea of continuing host–parasite coevolution, temporal variation in parasite infectivity changed little over time. In contrast, from the moment the parasite was first found in the sediments, we observed a steady increase in virulence over time, associated with higher fitness of the parasite.

It is notoriously difficult to study evolutionary dynamics in nature because time series over many generations are needed. Laminated lake sediments, however, provide a unique possibility of reconstructing the evolutionary dynamics of natural populations over time, because many aquatic organisms produce dormant propagule banks that archive gene pools. This approach has already revealed adaptive microevolution in natural populations of *Daphnia* in response to changes in predation pressure^{7,8} and anthropogenic perturbation⁹.

Although theory suggests that host–parasite interactions impose frequency-dependent selection, leading to Red Queen dynamics, empirical data on temporal dynamics in natural settings are entirely lacking^{3,6}. *Daphnia* and its microparasites show characteristics that are expected to lead to strong coevolutionary responses: there is ample genetic variation in *Daphnia* resistance to parasites¹⁰, host–parasite interactions are genotype specific^{11,12} and the genetic structure of *Daphnia* populations shifts during epidemics¹³. There is also evidence of local parasite adaptation¹⁴ and short-term parasite-mediated selection in *Daphnia*^{15,16}, which affects host–parasite dynamics¹⁷. Yet it has so far remained impossible to document the long-term dynamics of *Daphnia*–parasite coevolution in a natural setting⁶. The fact that both *Daphnia* and its parasites produce dormant propagule banks¹⁸, which accumulate in pond sediments, offers the opportunity to address this gap.

We sampled two sediment cores from a shallow pond (Oude Meren 2, ‘Abdij van ‘t Park’, Heverlee, Belgium¹⁸) where *Daphnia magna* coexists with the bacterial endoparasite *Pasteuria ramosa*. From different sediment layers of these cores we hatched *D. magna* clones from dormant eggs and picked up isolates of *P. ramosa*. Each depth represents a snapshot in the arms race of these antagonists, corresponding with a historical time fragment of the parasite and the *Daphnia* population. The oldest layer studied here (deepest depth 24 cm, a maximum of about 39 years old) represents the first time that both *D. magna* and *P. ramosa* co-occurred in this pond¹⁸. In two cross-infection experiments, we exposed *Daphnia* clones from eight (experiment 1) or seven (experiment 2) depths to parasite isolates from the next layer down, the same layer and the next layer up. Thus, the host was exposed to ‘past’, ‘contemporary’ and ‘future’ parasite isolates (further referred to as a time shift of parasites relative to host populations).

The reciprocal nature of Red Queen dynamics means that they are not always easily visible⁶, because the fitness of both antagonists may change only slightly over time³. Indeed, Red Queen dynamics may result in no or few apparent changes in the phenotypes of the interacting populations, even though there may be continuous changes in the underlying genetic structure. The Red Queen hypothesis has received much attention because it implies that the host benefits by producing genetically heterogeneous offspring by means of sexual reproduction, thus creating new defence mechanisms against fast-evolving parasites^{19,20}.

Our results show that over a timescale of about two to four years (the temporal resolution of our sediment layers), parasites tracked common host clones. On average, infectivity was higher when *Daphnia* were exposed to contemporary (average infectivity 0.65) parasites than to parasites from previous (average infectivity 0.55) growing seasons (Fig. 1). However, parasite adaptation was quickly lost, because average parasite infectivity was lower when *Daphnia* clones were confronted with future parasites (average infectivity 0.57) than with contemporary parasites (Fig. 1). There was a significant time-shift effect ($P = 0.05$) as well as a very significant depth \times time-shift interaction (Table 1). If past, contemporary and future parasites are designated P, C and F, respectively, the consistency of the pattern with the highest infectivity in the contemporary combinations across depths was high ($P < C$ in 10 of the 13 depths tested, and $C > F$ in 11) but was different in two depths (D5 and D7 from experiment 2; Fig. 1), generating the overall clone-depth \times time-shift interaction (Table 1). The time-shift effect remains significant when considering only periods in which multiple parasites were examined (experiment 1, $P < C$ in six of the seven depths tested, and

¹Laboratory of Aquatic Ecology and Evolutionary Biology, ²Laboratory for Animal Biodiversity and Systematics, Katholieke Universiteit Leuven, Charles de Bériotstraat 32, 3000 Leuven, Belgium. ³Interdisciplinary Research Center (IRC), Katholieke Universiteit Leuven, Campus Kortrijk, Etienne Sabbelaan 53, 8500 Kortrijk, Belgium. ⁴Universität Basel, Zoologisches Institut, Evolutionsbiologie, Vesalgasse 1, 4053 Basel, Switzerland. ⁵Institut National de la Recherche Agronomique (INRA), UR1282, Infectiologie Animale et Santé Publique, Nouzilly, F-37380, France.

*These authors contributed equally to this work.

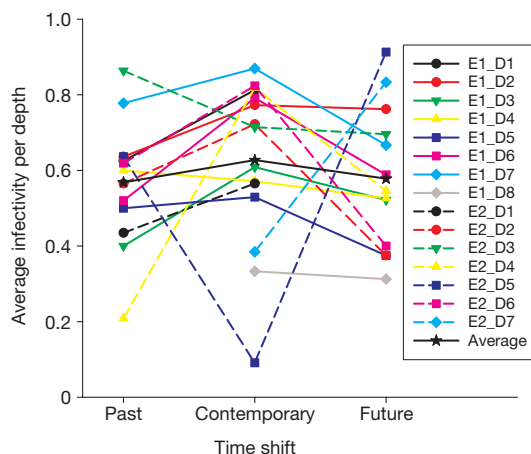


Figure 1 | Experimental results on temporal parasite adaptation. Average proportion of infected hosts when confronted with 'past', 'contemporary' and 'future' parasite isolates. Black stars, mean infectivity.

$C > F$ in seven; goodness of fit (deviation/degrees of freedom) = 1.293; time-shift effect $P = 0.01$; contrast analyses: $C - P$, odds ratio 1.869, $P = 0.0145$; $C - F$, odds ratio 1.71, $P = 0.036$. There was no significant time-shift effect when considering only periods in which single parasites were examined (experiment 2); instead, there is a very significant clone depth \times time-shift interaction in this experiment (experiment 2, $P < C$ and $C > F$ in four of the six depths tested; goodness of fit (deviation/degrees of freedom) = 1.206; interaction $P < 0.0001$), which we suspect reflects the large variation that occurs among parasite isolates and that is mis-attributed to a clone depth \times time-shift interaction when only one parasite is considered at each depth (Supplementary Information).

This rapid tracking of host genotypes by parasites is unlikely to be driven by strong continuous immigration of novel genotypes into the population (see Supplementary Information). We therefore conclude that our study reveals fast evolutionary changes, with parasites adapting to infect contemporary host genotypes. The loss of parasite infectivity in later years may be driven by parasites adapting to the changing host population. Often, when parasites adapt to different host genotypes, loss of adaptation to former hosts is seen²¹. These observations are consistent with a model of negative frequency-dependent selection. We tested the consistency of our experimental results with a haploid parasite and diploid host coevolutionary model based on a matching allele interaction matrix (see Supplementary Information). Simulations of this model resulted in negative frequency-dependent selection. Using realistic parameter settings for our host–parasite system, we found that parasite isolates from the near future were the most infective (Fig. 2a, b) and that hosts were least infected when exposed to parasites from the recent past. Allotemporary combinations farther into the future or the past showed a cyclic pattern of infectivity (Fig. 2a, b). This pattern changes when the measure of parasite infectivity is averaged over several generations, as in our experiment, in which hosts and parasites were pooled over 2-cm sediment layers (which may contain 10–20

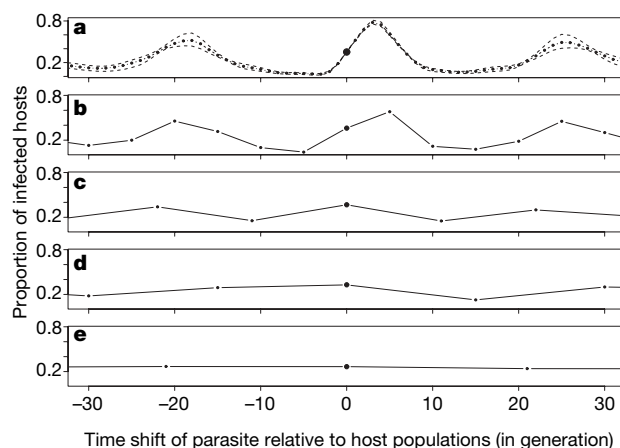


Figure 2 | Parasite infectivity as a function of the time shift of the parasites relative to the host population in the model simulation. Parasite infectivity averaged over 1 (a) 5 (b), 11 (c), 15 (d) and 21 (e) generations. A negative time refers to parasites from the past. Scatters around the line represent the 95% confidence interval. The compatibility matrix of parasite and host genotypes is defined by a matching allele model (MAM) with one locus and four alleles (Supplementary Information).

Daphnia generations). This yields a pattern in which contemporary parasites are more infective than past and future parasite isolates, which is consistent with our results (Fig. 2c). As expected, any further averaging causes a loss of dynamics (Fig. 2d, e). The dynamics seen in the model are qualitatively robust to various parameter settings: with more alleles the peaks spread out, whereas they become flatter when virulence is reduced (Supplementary Figs 2 and 3). Higher mutation rates tend to promote faster adaptation and result in smoother adaptation dynamics (Supplementary Fig. 4).

Our results are consistent with studies on mechanistic aspects of coevolution in the same host–parasite system^{10,11,22}. Furthermore, it has been shown that *Pasteuria* epidemics can select for resistance in *Daphnia*²³ and that genetic variation as observed under laboratory conditions explains disease occurrence under natural conditions²⁴. Our model, which was streamlined to match the details of our experimental design, indicates that our results are consistent with antagonistic coevolution on the basis of a matching-allele-type assumption. However, the model also shows that temporal adaptation differs strongly in strength and in cycle length when certain aspects of the host–parasite system are altered. Such changes are crucial in relation to the experimental design, because different time shifts or thicknesses of sediment layers (see Fig. 2 and Supplementary Fig. 2) could make it more or less likely that the pattern seen in our study will be found (Fig. 1). Indeed, an earlier one-year time-shift study employing a different design and working at a different location was unable to find evidence for coevolution²⁵. Furthermore, the arms race may be disguised altogether under more complex ecological conditions, for example, where other antagonists, such as predators, have a dominant impact²⁶. Our analysis also reveals that some of the contrasts between past and contemporary time layers and between contemporary and future layers deviate from the average picture (Fig. 1 and Supplementary Fig. 5). At certain times, one or the other antagonist may lag in the arms race. All these factors highlight that the signature of antagonistic coevolution in natural populations may be found only in host–parasite systems with certain properties and may need experimental conditions well chosen to match the biology of the system.

The continuous time series from the sediment cores did not reveal a net change in parasite infectivity over time (Fig. 3a, d). However, we found evidence for gradual changes in other components of parasite fitness. Starting from the first record of the parasite (deepest depth 24 cm), production of the *Pasteuria* transmission stages increased with isolates of more recent origin (Fig. 3b, e). Increased parasite

Table 1 | Results of logistic regression on infectivity

Parameter	Degrees of freedom	Residuals of deviance	Residuals of degrees of freedom	Deviance	P
Null	–	–	844	1,143.56	–
Time shift	2	5.89	842	1,137.67	0.05
Experiment	1	0.20	841	1,137.47	0.65
Clone depth	7	11.61	834	1,125.86	0.11
Clone depth \times time shift	12	57.44	820	1,063.11	<0.0001

A binomial model, logit link, was used, with time shift reflecting the difference between the combinations of host clone and parasite isolate depth (three levels: 'past', 'contemporary' and 'future'), excluding non-significant interactions.

spore production was associated with stronger fecundity reduction in the *Daphnia* host (product moment correlation between spore production and fecundity reduction: $n = 15$, $R = 0.98$, $P < 0.0001$; see also ref. 27). Thus, the parasite decreased host fecundity progressively over time (Fig. 3c, f).

This increase in spore production and virulence in *Pasteuria* over time (but not its infectivity; Fig. 3) may reflect adaptation of the parasite to the host. It is unlikely to be explained by differences in the duration of dormancy of the parasite spores (as a result of genetic deterioration or trade-offs between dormancy and other traits, for example). All parasite spores were revived by a growth cycle in *Daphnia* hosts after they were picked up from the sediment. Consequently, all spores used in the experiment were physiologically of the same age. Moreover, *Pasteuria* that were randomly isolated with a standard set of host clones did not show an increase in spore production over time¹⁸ (results not shown). Furthermore, it is unlikely that endospores such as those from *Pasteuria* accumulate heritable damage that is reflected in fitness parameters within the time span covered in our sediment core (oldest layer 24 cm). Bacilli, which are closely related to *Pasteuria*, protect the DNA of their endospores with small acid-soluble proteins during dormancy²⁸. An absence of genetic degeneration during dormancy has been reported in plant seeds up to 150 years old²⁹. At present it is not clear what mechanism can explain the different dynamics in parasite spore production and virulence compared with infectivity.

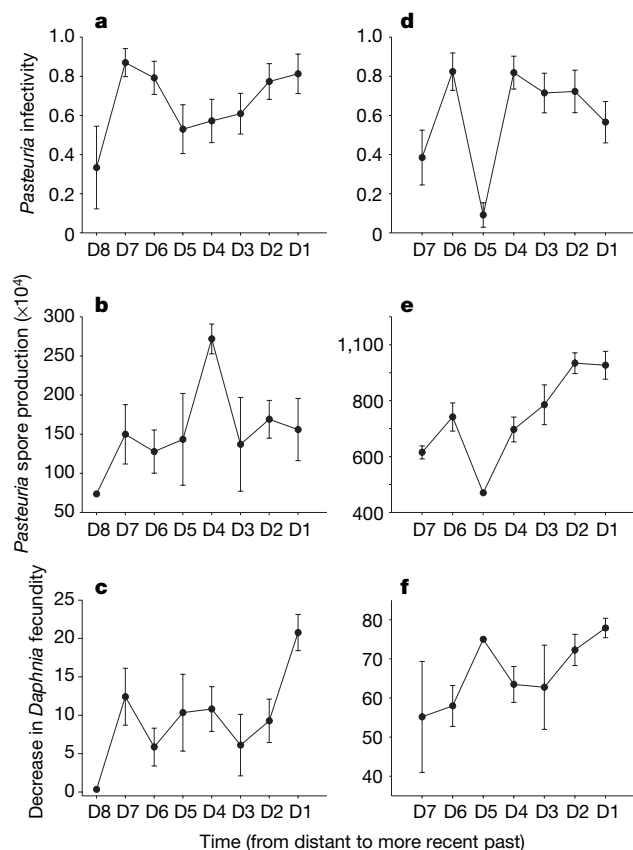


Figure 3 | *Pasteuria* infectivity and virulence over time. a–c, Experiment 1; d–f, experiment 2. a, d, *Pasteuria* infectivity; b, e, *Pasteuria* spore production; c, f, decrease in *Daphnia* fecundity (clonal average of control individuals minus clonal average of infected individuals). Results are means \pm s.e.m. and include only ‘contemporary’ combinations. Time reflects parasite isolate depth, with D8 referring to the oldest sediment layer. The effects of parasite isolate depth as continuous covariable in a general linear model (including significant effects of experiment; parasite isolate depth was nested in experiment) are as follows: on *Pasteuria* infectivity, $F_{2,108} = 0.52$, $P = 0.59$; on *Pasteuria* spore production, $F_{2,80} = 9.6$, $P = 0.0002$; on decrease in *Daphnia* fecundity, $F_{2,81} = 5.03$, $P = 0.009$.

It is likely that different genes contribute to different parasite fitness components and thus follow different evolutionary dynamics.

In this historical reconstruction of host–parasite interactions in a natural population, we provide empirical evidence for coevolution based on fluctuating selection. Such Red Queen dynamics have been suggested to have a role in widespread biological phenomena such as the evolution and maintenance of sexual reproduction^{19,20} and the maintenance of genetic polymorphism at disease loci³⁰. In this sense, antagonistic coevolution may have a key role in the evolution of biodiversity on Earth².

METHODS SUMMARY

Daphnia host clones and parasite isolates were derived from two sediment cores (one per experiment) that were sliced into depth increments. From each core, eight 2-cm depth increments were used, covering a time interval of about 17 to 28 years—thus, about 2–4 years per 2-cm depth—with the age of the deepest sediment layer being about 29–39 years. On the basis of the similar profile of the number of dormant eggs in the different depths of the cores of both experiments (Supplementary Fig. 1), we numbered the depths used equally between the cores. Host clones were obtained through hatching of dormant eggs from each depth. Parasite isolates were obtained by exposing a random set of *Daphnia* clones to sediment from each depth. For experiment 1, we obtained three parasite isolates per depth (eight depths \times 3 = 24 in total). For experiment 2, we obtained only one parasite isolate per depth from the first seven depths. After controlling for maternal effects, neonates of the host clones were isolated and individually exposed to a parasite or a control treatment. In the parasite treatments, the host clones were exposed to ‘past’, ‘contemporary’ and ‘future’ parasite populations (referred to as time-shift treatment). Therefore, for each depth, three (experiment 1) or eight (experiment 2) host clones were exposed to three (experiment 1) or one (experiment 2) parasite isolate(s) from the next layer down, the same layer and the next layer up, in three replicates. During the experiments, the number of juveniles produced and the time to castration (when *Pasteuria* prevents *Daphnia* from reproducing) was recorded. After 26 days (experiment 1) or 23 days (experiment 2), all *Daphnia* were checked for infection and *Pasteuria* spore production was recorded.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 June; accepted 21 September 2007.

Published online 14 November 2007.

- Harvell, D. Ecology and evolution of host–pathogen interactions in nature. *Am. Nat.* **164**, S1–S5 (2004).
- Thompson, J. N. & Cunningham, B. M. Geographic structure and dynamics of coevolutionary selection. *Nature* **417**, 735–738 (2002).
- Woolhouse, M. E. J., Webster, J. P., Domingo, E., Charlesworth, B. & Levin, B. R. Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nature Genet.* **32**, 569–577 (2002).
- Fenner, F. & Fantini, B. *Biological Control of Vertebrate Pests* (CABI Publishing, Wallingford, 1999).
- Lively, C. M. & Dybdahl, M. F. Parasite adaptation to locally common genotypes. *Nature* **405**, 679–681 (2000).
- Little, T. J. The evolutionary significance of parasitism: do parasite-driven genetic dynamics occur *ex silico*? *J. Evol. Biol.* **15**, 1–9 (2002).
- Cousyn, C. *et al.* Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of natural genetic changes. *Proc. Natl Acad. Sci. USA* **98**, 6256–6260 (2001).
- Kerfoot, W. C. & Weider, L. J. Experimental paleoecology (resurrection ecology): Chasing Van Valen’s Red Queen hypothesis. *Limnol. Oceanogr.* **49**, 1300–1316 (2004).
- Hairton, N. G. *et al.* Rapid evolution revealed by dormant eggs. *Nature* **401**, 446 (1999).
- Little, T. J. & Ebert, D. Associations between parasitism and host genotype in natural populations of *Daphnia* (Crustacea: Cladocera). *J. Anim. Ecol.* **68**, 134–149 (1999).
- Carius, H. J., Little, T. J. & Ebert, D. Genetic variation in a host–parasite association. *Evolution Int. J. Org. Evolution* **55**, 1146–1152 (2001).
- Decaestecker, E., Vergote, A., Ebert, D. & De Meester, L. Evidence for strong host clone–parasite species interactions in the *Daphnia* microparasite system. *Evolution Int. J. Org. Evolution* **57**, 784–792 (2003).
- Mitchell, S. E., Read, A. F. & Little, T. J. The effect of a pathogen epidemic on the genetic structure and reproductive strategy of the crustacean *Daphnia magna*. *Ecol. Lett.* **7**, 848–858 (2004).
- Ebert, D. Virulence and local adaptation of a horizontally transmitted parasite. *Science* **265**, 1084–1086 (1994).

15. Capaul, M. & Ebert, D. Parasite-mediated selection in experimental *Daphnia magna* populations. *Evolution Int. J. Org. Evolution* **57**, 249–260 (2003).
16. Haag, C. R. & Ebert, D. Parasite-mediated selection in experimental metapopulations of *Daphnia magna*. *Proc. R. Soc. Lond. B* **271**, 2149–2155 (2004).
17. Duffy, M. A. & Sivers-Becker, L. Rapid evolution and ecological host–parasite dynamics. *Ecol. Lett.* **10**, 44–53 (2007).
18. Decaestecker, E., Lefever, C., De Meester, L. & Ebert, D. Haunted by the past: evidence for dormant stage banks and epibionts of *Daphnia*. *Limnol. Oceanogr.* **49**, 1355–1364 (2004).
19. Hamilton, W. D. Sex versus non-sex parasite. *Oikos* **35**, 282–290 (1980).
20. West, S. A., Lively, C. M. & Read, A. F. A pluralist approach to sex and recombination. *J. Evol. Biol.* **12**, 1003–1012 (1999).
21. Ebert, D. Experimental evolution of parasites. *Science* **282**, 1432–1435 (1998).
22. Little, T. J., Watt, K. & Ebert, D. Parasite–host specificity: experimental studies on the basis of adaptation. *Evolution Int. J. Org. Evolution* **60**, 31–38 (2006).
23. Duncan, A. & Little, T. J. Parasite-driven genetic change in a natural population of *Daphnia*. *Evolution Int. J. Org. Evolution* **61**, 796–803 (2007).
24. Little, T. J. & Ebert, D. The cause of parasitic infection in natural populations of *Daphnia* (Crustacea: Cladocera): the role of host genetics. *Proc. R. Soc. Lond. B* **267**, 2037–2042 (2000).
25. Little, T. J. & Ebert, D. Temporal patterns of genetic variation for resistance and infectivity in a *Daphnia*–microparasite system. *Evolution Int. J. Org. Evolution* **55**, 1146–1152 (2001).
26. Decaestecker, E., De Meester, L. & Ebert, D. In deep trouble: habitat selection constrained by multiple enemies in zooplankton. *Proc. Natl Acad. Sci. USA* **99**, 5481–5485 (2002).
27. Ebert, D., Carius, H. J., Little, T. J. & Decaestecker, E. The evolution of virulence when parasites cause host castration and gigantism. *Am. Nat.* **164S**, S19–S32 (2004).
28. Kosman, J. & Setlow, P. Effects of carboxy-terminal modifications and pH on binding of a *Bacillus subtilis* small acid soluble spore protein to DNA. *J. Bacteriol.* **185**, 6095–6103 (2003).
29. Vavrek, M. C., Mccraw, J. B. & Bennington, C. C. Ecological genetic variation in seed banks. III. Phenotypic and genetic differences between young and old seed populations of *Carex bigelowii*. *J. Ecol.* **79**, 645–662 (1991).
30. Wegner, K. M., Kalbe, M., Kurtz, J., Reusch, T. B. H. & Milinski, M. Parasite selection for immunogenetic optimality. *Science* **301**, 1343 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. L. Burks, B. Jansen, T. Huyse, T. J. Little, J. Mergeay, D. J. Mikolajewski, R. Ortells, F. Van de Meutter, J. Vanoverbeke and G. Verbeke for discussion and advice; J. Vandekerckhove, D. Verreydt, H. Michels and A. Wollebrants for technical help; and S. Sweizig for linguistic corrections. We acknowledge constructive comments by A. Read and Y. Michalakis. This research was supported by grants from the KULeuven Research Fund to E.D., the Fonds voor Wetenschappelijk Onderzoek – Vlaanderen (F.W.O) to E.D. and J.A.M.R., the KULeuven Research Fund to L.D.M., INRA and the Swiss Nationalfonds to S.G., and the Swiss Nationalfonds to D.E.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.D. (ellen.decaestecker@kuleuven-kortrijk.be).

METHODS

Sediment cores. For experiment 1, a core with a diameter of 5 cm was taken in October 2002 and was sliced in depth increments of 2 cm. For experiment 2, a core with a diameter of 15 cm was taken in January 2006 and sliced in increments of 1 cm. The outer 1 mm of each slice was removed. Dating of the core was performed on the basis of the assumption of constant sedimentation rates in terms of sediment mass (see refs 31, 32). Because the time axis should be considered indicative, only rough statements with respect to time can be made.

Daphnia clones and parasite isolates. Host clones were collected by hatching dormant eggs that were isolated from ephippia. Hatching was induced at 15 °C under a 16 h:8 h light/dark cycle. In all further experiments, *Daphnia* clones were kept under standardized culture conditions (19 ± 1 °C and a light/dark cycle of 16 h:8 h) and fed daily with the alga *Scenedesmus acuminatus*.

The spores of the parasite isolates of the cross-infection experiments were isolated by exposing a random set of *Daphnia* clones (16 clones in experiment 1 and 20 clones in experiment 2, hatched from all sediment depths) to sediment (including *P. ramosa* spores) from each depth separately. Two replicates of each clone were individually exposed to an equal amount of sediment from all depths in 50-ml jars (4×10^4 algal cells ml⁻¹ per day). After 28 days of exposure all infected *Daphnia* were isolated for use as a source of spores for the parasite isolates of the cross-infection experiment. For each sediment depth, each parasite isolate was isolated from a different clone. To obtain a sufficient number of spores from each parasite isolate in the cross-infection experiments, all parasite spores were propagated once in *Daphnia* juveniles from the same clone as the *Daphnia* mother from which the spores were originally isolated. Spore solutions, obtained from one 'squashed' infected mother, were divided between two (experiment 1) or three (experiment 2) 50-ml experimental jars with juveniles. For experiment 1, we obtained 24 parasite isolates in total, three from each 2-cm depth (eight depths \times 3 = 24). In experiment 2, we lost several isolates during the propagation phase and finally obtained only seven parasite isolates in total, one from each of the first seven 2-cm depths. For each parasite isolate, infected *Daphnia* were then isolated, pooled and ground up to prepare final spore solutions for the cross-infection experiment. Parasite isolates from each depth were independent from the host clones used in the cross-infection experiment because, in general, the *Daphnia* clones that picked up the parasite spores originated from a different depth from the host clones in the cross-infection experiment. Spore solutions and the amount of added *Daphnia* tissue were standardized across all parasite isolates in each final cross-infection experiment.

Cross-infection experiments. Maternal lines were started up with adult female *Daphnia* from clonal stock cultures kept for several generations in the laboratory and derived from a single dormant egg. Neonates of the second clutch of the second-generation maternal lines were isolated and individually exposed to a parasite or control treatment in 100-ml (experiment 1) or 50-ml (experiment 2) jars. In experiment 1, whenever possible, three *Daphnia* clones (other than those used to pick up the parasite isolates) from each 2-cm depth were exposed to nine parasite isolates (three parasite isolates from the next layer down, the same layer and the next layer up) in three replicates. Only two parasite isolates were used for

depth five and only one for depth eight. In all parasite treatments, 2.5×10^5 spores ml⁻¹ were added on day one, and 2.5×10^4 spores ml⁻¹ on day seven. For experiment 2, we did not obtain as large numbers of spores as in the first experiment, and we used a lower spore concentration in this experiment. In experiment 2, whenever possible, eight *Daphnia* clones from each 2-cm depth were exposed to three parasite isolates (one parasite isolate from the next layer down, the same layer and the next layer up) in three replicates. We did not obtain parasite isolates from the lowest depth in experiment 2; E2_D8 was therefore not included in the analyses. In all parasite treatments, 1.3×10^3 spores ml⁻¹ were added on day 1, and 0.65×10^3 spores ml⁻¹ were added on day 3. In experiment 1, all *Daphnia* were fed daily with 0.8×10^5 algal cells ml⁻¹; in experiment 2, they were fed with 0.5×10^5 algal cells ml⁻¹ for the first six days, and afterwards with 1.5×10^5 algal cells ml⁻¹. Food levels were relatively equal (difference of 0.3×10^5 algal cells ml⁻¹) and low in both experiments in the first days of the experiment, to ensure that parasite uptake would be high. In comparison with experiment 1, food quantity in the second part of experiment 2 was considerably higher. This was to ensure that our previous patterns could also be found when the *Daphnia* were not stressed. The medium was refreshed every time the host released a clutch, or at least every fourth day.

Data analysis. In the infectivity analysis we were interested in the effect of host clones and the origin of parasite isolates. We tested whether infectivity in contemporary associations was higher than in hosts exposed to future or past parasites. We performed a logistic regression on infectivity in R³³: binomial model, logit link. Factors were time shift (time difference between the different depth combinations past, contemporary and future), experiment and clone depth (nested in experiment). Non-significant interactions were excluded. When we included host clone in the model as a random effect, our results on time shift did not differ.

In the virulence analysis we were interested in whether there was an evolutionary change in *Pasteuria* spore production and virulence (estimated as parasite-induced decrease in host fecundity and earlier time to castration (when *Pasteuria* prevents *Daphnia* from reproducing)) over time. We only included contemporary associations. General linear models with time shift and experiment as categorical factors and parasite isolate depth (nested in experiment) as a continuous factor (there were no parasite isolate effects) were performed on the averages for each combination of host clone and parasite isolate. Infection rate was included as a continuous covariable in the general linear models of the *Pasteuria* spore production and virulence parameters. There was no interaction between parasite isolate depth and time shift in any of the parameters tested.

31. Cousyn, C. et al. Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of natural genetic changes. *Proc. Natl Acad. Sci. USA* **98**, 6256–6260 (2001).
32. Decaestecker, E., Lefever, C., De Meester, L. & Ebert, D. Haunted by the past: evidence for dormant stage banks and epibionts of *Daphnia*. *Limnol. Oceanogr.* **49**, 1355–1364 (2004).
33. R Development Core Team. The R project for statistical computing (<http://www.R-project.org>) (2005).

LETTERS

Methanotrophy below pH 1 by a new *Verrucomicrobia* species

Arjan Pol¹, Klaas Heijmans¹, Harry R. Harhangi¹, Dario Tedesco², Mike S. M. Jetten¹ & Huub J. M. Op den Camp¹

Mud volcanoes, mudpots and fumaroles are remarkable geological features characterized by the emission of gas, water and/or semi-liquid mud matrices¹ with significant methane fluxes to the atmosphere (10^{-1} to 10^3 t y^{-1})^{2–4}. Environmental conditions in these areas vary from ambient temperature and neutral pH to high temperatures and low pH. Although there are strong indications for biological methane consumption in mud volcanoes^{4,5}, no methanotrophic bacteria are known that would thrive in the hostile conditions of fumaroles (temperatures up to 70 °C and pH down to 1.8)². The first step in aerobic methane oxidation is performed by a soluble or membrane-bound methane mono-oxygenase. Here we report that *pmoA* (encoding the β -subunit of membrane-bound methane mono-oxygenase) clone libraries, made by using DNA extracted from the Solfatara volcano mudpot and surrounding bare soil near the fumaroles, showed clusters of novel and distant *pmoA* genes. After methanotrophic enrichment at 50 °C and pH 2.0 the most distant cluster, sharing less than 50% identity with any other described *pmoA* gene, was represented in the culture. Finally we isolated an acidiphilic methanotrophic bacterium *Acidimethylosilex fumarolicum* SolV belonging to the Planctomycetes/Verrucomicrobia/Chlamydiae superphylum⁶, 'outside' the subphyla of the Alpha- and Gammaproteobacteria containing the established methanotrophs. This bacterium grows under oxygen limitation on methane as the sole source of energy, down to pH 0.8—far below the pH optimum of any previously described methanotroph. *A. fumarolicum* SolV has three different *pmoA* genes, with two that are very similar to sequences retrieved from the mudpot. Highly homologous environmental 16S rRNA gene sequences from Yellowstone Park show that this new type of methanotrophic bacteria may be a common inhabitant of extreme environments. This is the first time that a representative of the widely distributed Verrucomicrobia phylum, of which most members remain uncultivated⁶, is coupled to a geochemically relevant reaction.

Significant amounts of geological methane, produced within the Earth's crust, are currently released naturally into the atmosphere^{3,7,8}. The preliminary global estimate of these methane emissions indicates that there are probably more than enough sources to provide the amount of methane required to account for the suspected missing source of global methane⁸. Recent findings from the Haakon Mosby and Carpatian mud volcanoes showed that these systems may also act as sinks for this geological methane^{4,5,9}. At these sites with moderate environmental conditions (2–25 °C and a neutral pH), 16S rRNA genes of both aerobic and anaerobic methane-oxidizing microorganisms were present. In contrast, fumaroles such as those located in the Solfatara at Pozzuoli near Naples (southern Italy), which also emit significant amounts of methane (73 tonnes of CH_4 per km^2 per year)², are characterized by soils with a low pH (down to 1.0) and elevated temperatures (up to 70 °C). The H_2S -rich sulphurous fumes

at these sites are microbially oxidized into sulphuric acid, creating an extremely acidic environment. The very acidic soil of the Solfatara was shown to support significant methane consumption², but so far it is unknown which microbes could be responsible for this consumption. Obligately aerobic methanotrophs are assumed to be a unique group of bacteria, belonging to either the Alpha or Gamma subclass of the Proteobacteria, which use methane as the sole source of energy and carbon¹⁰. So far, all aerobic methanotrophs have been shown to contain a membrane-bound particulate methane mono-oxygenase (pMMO), except for *Methylocella* sp. that was reported to have only the soluble, cytoplasmic form of MMO (sMMO)¹¹. The *pmoA* gene (encoding the 24 kDa β -subunit of this membrane bound MMO¹²) is generally used as a phylogenetic marker for methanotrophic bacteria. Methanotrophs are widespread in nature and are mostly neutrophilic and mesophilic. However, on the basis of molecular surveys, in the last decade isolation and characterization of more extremophilic proteobacterial methanotrophs was initiated¹³. Thus far, the lowest pH values still supporting methanotrophic activity were reported for bacteria isolated from peat bogs^{11,14,15}. These bacteria belong to genera of the Alpha subclass of the Proteobacteria (*Methylocella*, *Methylocapsa* and *Methylocystis*), and showed growth between pH 4.2 and 7.5 with a maximum methane-oxidizing activity around pH 5.0.

The inner part of the Solfatara, characterized by a central mudpool (fangaia) surrounded by bare, acid soil (pH 1–2), was sampled and DNA was extracted to start a molecular survey of *pmoA* genes. Here we report the presence of *pmoA* genes in an environmental clone library constructed using this DNA as a PCR template for the widely applied *pmoA* primer set A189/A682 (ref. 16), which also may amplify the gene of ammonium mono-oxygenase β subunit (*amoA*). We were only able to amplify *pmoA* genes using non-restrictive conditions (annealing temperature lowered from 56 °C to 48 °C; no false-positive clones obtained), pointing to the presence of *pmoA* genes with low similarity to known sequences. This is supported by the phylogenetic analyses of the *pmoA* sequences, which show that the Solfatara *pmoA* sequences group into two clusters: one represents a completely new, deep branch within the *pmoA/amoA* phylogenetic tree, sharing very low homology to known sequences (Fig. 1); the other cluster groups with the Gammaproteobacterial methanotrophs.

Intrigued by the new *pmoA* sequences, we used mud and mixed-soil samples from this site to start enrichment cultures at 50 °C and pH 2 with methane as the sole source of energy and carbon. After 3 weeks, methane consumption was observed in both soil and mud incubations. Non-restrictive PCR amplification of *pmoA* sequences, with DNA from the enrichment as a template, resulted in five clones (from two different enrichments) with sequences almost identical to the distant group within the environmental clones (Fig. 1). Repeated

¹Department of Microbiology, IWW, Radboud University Nijmegen, Toernooiveld 1, NL-6525 ED Nijmegen, The Netherlands. ²Dipartimento di Scienze Ambientali, Seconda Università di Napoli, Via Vivaldi 43, 81100 Caserta, Italy.

serial transfers of the mud culture into fresh medium (see Methods) and finally diluting the culture onto floating polycarbonate filters¹⁷ resulted in a pure culture, named strain SolV. Tiny whitish colonies appeared on the filters after 1 week and microscopic observation revealed only one rod-shaped morphotype. When exponentially growing cells of SolV were tested, the sMMO activity test (conversion of naphthalene to naphthol) was negative, but pMMO activity (particulate MMO; using propylene) could easily be measured (50 nmol per min per mg of protein). Genomic DNA from SolV was extracted and subjected to pyrosequencing¹⁸. From these data, we could identify many genes of C1 metabolism (Table 1), indicating that strain SolV may use a new combination of the serine, tetrahydrofolate and ribulose-1,5 biphosphosphate pathways for carbon assimilation. The diagnostic genes of the ribulose-monophosphate pathway seem to be absent (Table 1). Conversion of formaldehyde seems to be mediated by a tetrahydrofolate-dependent pathway or directly by formaldehyde dehydrogenase (activity 110 nmol per min per mg of protein). The methanol dehydrogenase activity was 60 nmol per min per mg of protein and the *mxoF* gene showed 50% identity to *mxoF* of *Methylococcus capsulatus*. None of the subunits of sMMO was found. However, two complete *pmoCAB* operons and one *pmoCAB* cluster with a partial *pmoC* were identified. Several (two to nine) mismatches with *pmoA* primers A189/A682 were found (Supplementary Fig. 1), explaining the low recovery in PCR amplification from environmental DNA. However, all signature

amino acids of PmoA were present, whereas the signature amino acids of AmoA were absent¹⁹. Of all 42 highly conserved amino acids in all bacterial PmoA/AmoA proteins¹⁹, 6 to 8 were not shared by one or more of the *pmoA* genes from strain SolV (Supplementary Fig. 2). Phylogenetic analysis of the *pmoA* genes showed that *pmoA1* and *pmoA2* are highly similar to the environmental sequences from the Solfatara and the enrichments (Fig. 1, and see above). The *pmoA3* gene represents another completely new, deep branch. Together these three new *pmoA* sequences indicate that methanotrophic bacteria are phylogenetically much more diverse than currently assumed. Recent genomic data have shown that two either identical or distantly related *pmoA* genes can be present in one Alpha- or Gamma-proteobacterial methanotroph^{20–22}. Expression of *pmoA1* and *pmoA2* messenger RNA was confirmed by RT-PCR on mRNA extracted from methane-grown SolV cells using specific primers (see Methods). The stacked membrane structures characteristic for methanotrophs expressing pMMO were not observed in SolV by transmission electron microscopy (Supplementary Fig. 3). Instead, circular bodies of about 50–70 nm were observed after fixation with glutaraldehyde or cryofixation. These bodies may be reminiscent of the vesicles observed in the acidiphilic methanotroph *Methylocella palustris*²³.

Growth of strain SolV occurred between pH 0.8 and 5.8 (Fig. 2). The temperature optimum is 55 °C, with only minor growth observed below 40 °C and above 65 °C. The maximum-specific-growth rate on methane was 0.07 h⁻¹ (doubling time 10 h). Carbon dioxide and the

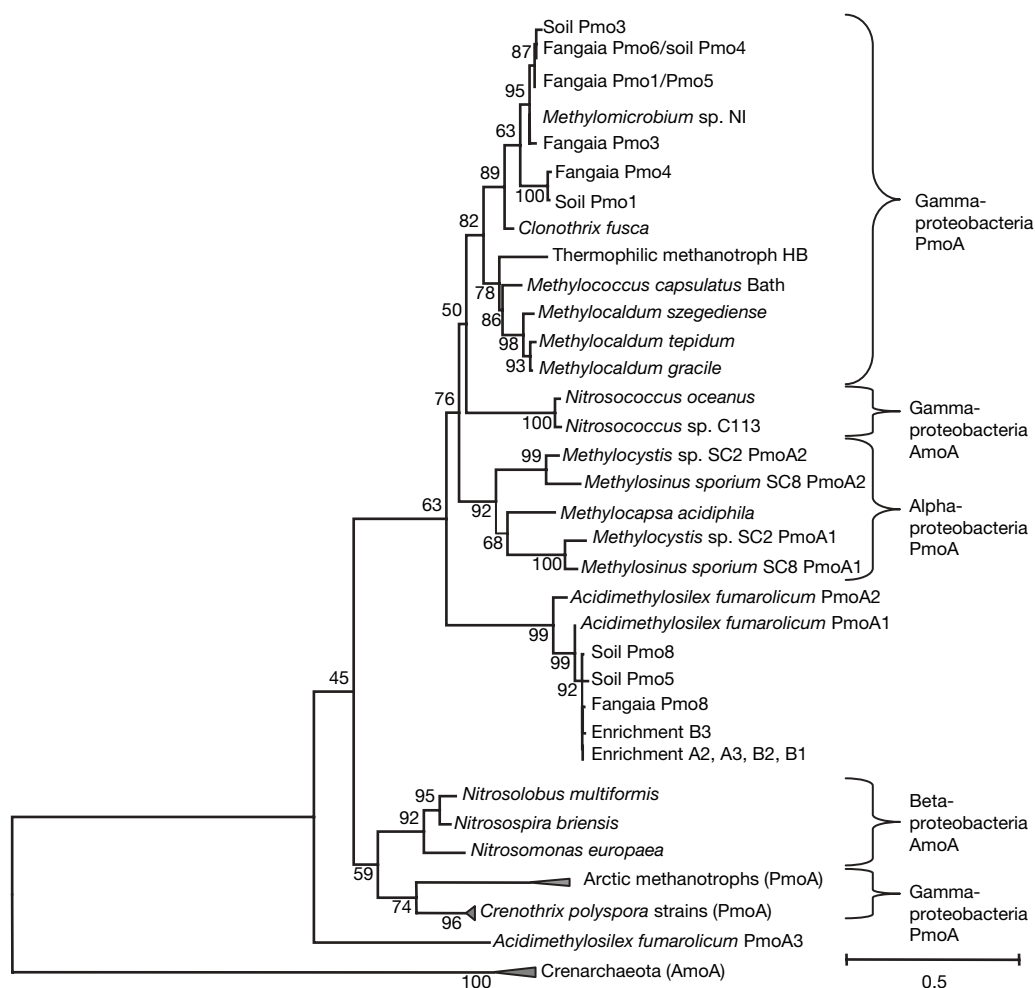


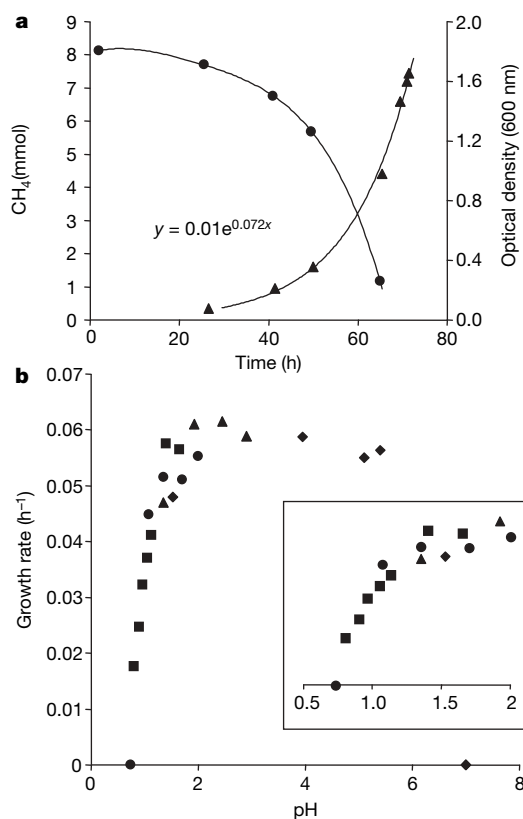
Figure 1 | Phylogenetic relationship among deduced PmoA and AmoA proteins. The neighbour-joining tree calculated with the PAM Dayhoff matrix is shown with bootstraps values of 500 replicates at the branches. The bar represents a 50% estimated-sequence divergence. Application of different methods of compiling trees revealed congruent tree topologies. The

fangaia Pmo and soil Pmo prefix refer to environmental clones from DNA extracted from the central mudpot and bare, acid soil, respectively. Enrichment refers to clones obtained from DNA extracted from two different enrichments (A and B).

Table 1 | Genes of C1 metabolism of *Acidimethylosilex fumarolicum* SolV

Enzyme	Enzyme Commission (EC) number	Gene	BLASTP search against <i>Methylococcus capsulatus</i>			
			Identity (%)	Similarity (%)	Expected (E)-value	GenBank
Methane mono-oxygenase	1.14.13.25	<i>pmoA1</i>	53	71	4.6×10^{-80}	mca1797
		<i>pmoA2</i>	57	74	5.5×10^{-81}	mca1797
		<i>pmoA3</i>	41	62	1.2×10^{-51}	mca1797
		<i>pmoB1</i>	39	57	6.8×10^{-75}	mca2853
		<i>pmoB2</i>	40	58	1.6×10^{-76}	mca2853
		<i>pmoB3</i>	38	56	8.7×10^{-75}	mca2853
		<i>pmoC1*</i>				
		<i>pmoC2</i>	58	72	5.9×10^{-77}	mca0295
		<i>pmoC3</i>	43	60	8.0×10^{-50}	mca0295
		<i>mmoX</i>	Not present in SolV			
Methanol dehydrogenase	1.1.99.8	<i>mxoF</i>	50	64	2.4×10^{-169}	mca0299
		<i>mxoJ</i>	36	55	1.7×10^{-36}	mca0300
		<i>mxoG</i>	34	51	2.0×10^{-8}	mca0781
		<i>adhP</i>	41	58	8.5×10^{-68}	mca0775
Formaldehyde dehydrogenase	1.2.99.3	<i>fae</i>	Not present in SolV			
Formaldehyde-activating enzyme	4.3.-.-	<i>fdhA</i>	50	67	3.1×10^{-28}	mca1393
Formate dehydrogenase		<i>fdhB</i>	62	78	6.7×10^{-172}	mca1392
		<i>fdhC</i>	68	81	0	mca1391
		<i>fdhD</i>	48	67	3.0×10^{-12}	mca1389
Serine-glyoxylate aminotransferase	2.6.1.45	<i>agxt/spt</i>	31	50	6.1×10^{-40}	mca1406
Hydroxypyruvate dehydrogenase	1.1.1.29	<i>hprA</i>	32	52	5.0×10^{-23}	mca1407
Formate-tetrahydrofolate ligase	6.3.4.3	<i>fhs</i>	53	70	1.5×10^{-165}	mca2219
Serinehydroxymethyl transferase	2.1.2.1	<i>glyA</i>	57	75	2.6×10^{-135}	mca1660
5-formyltetrahydrofolate cycloligase	6.3.3.2	<i>mthfs</i>	29	45	5.7×10^{-11}	mca2773
Methylenetetrahydrofolate dehydrogenase/ methenyltetrahydrofolate cyclohydrolase	1.5.1.5/3.5.4.9	<i>folD</i>	Not present in <i>M. capsulatus</i> †			
Hexulose-6-phosphate synthase	4.1.2.-	<i>hspA</i>	Not present in SolV			
Hexulose-6-phosphate isomerase	5.-.-.-	<i>sgbU</i>	Not present in SolV ‡			
Ribulose biphosphate carboxylase	4.1.1.39	<i>cbbL</i>	60	75	4.9×10^{-165}	mca2743
		<i>cbbS</i>	41	62	4.5×10^{-18}	mca2744
Phosphoribulokinase	2.7.1.19	<i>cbbP</i>	64	80	6.6×10^{-108}	mca3051
Phosphoglycerate kinase	2.7.2.3	<i>cbbK</i>	40	61	1.5×10^{-76}	mca2021
Glyceraldehyde-3-phosphate dehydrogenase	1.2.1.13	<i>cbbG</i>	49	64	1.4×10^{-72}	mca2598

Genes of C1 metabolism were identified in an assembly of the genome after pyrosequencing. The assembly was produced from 88.9 Mb of sequence information and resulted in a 35-fold coverage, based on an estimated genome size of 2.5 Mb. Translated protein sequences, based on genes identified, were used for a BLAST search in the *Methylococcus capsulatus* genome (<http://pedant.gsf.de/>). *partial gene (48 amino acids); †best NCBI BLAST hit with *folD* from *Prosthecochloris aestuarii* (identity 51%; similarity 71%; E-value 1.0×10^{-73}); ‡best NCBI BLAST hit with *gutQ* (sugar phosphate isomerase family) from *Burkholderia phytofirmans*, (identity 45%; similarity 66%; E-value 2.0×10^{-78}).



inorganic fraction of mud water stimulated growth. Methane was converted to carbon dioxide according to a stoichiometry that is typical for methanotrophs: $\text{CH}_4 + 1.6 \text{O}_2 \rightarrow 0.65 \text{CO}_2 + 1.55 \text{H}_2\text{O} + 0.35 \text{CH}_2\text{O}$ (biomass) with a yield of 6.4 g of dry weight per mol of methane.

Acetate, malate, succinate, formate, formaldehyde and yeast extract (all at 1 g l^{-1}) completely inhibited growth of SolV on methane at pH 2. The bacterium apparently is very sensitive towards uncoupling by small organic acids at low pH values, because at pH 5 formate ($\text{pK}_a 3.75$) did not inhibit growth. No growth took place above 100 mM NaCl or in media containing glucose. In addition to methane, hydrogen gas was also oxidized. Strain SolV grew well on methanol, but the added methanol completely repressed methane consumption. After methanol was depleted, methane consumption and growth started only after 4 h. Ethane inhibited growth although it was converted simultaneously with methane as a competitive substrate at virtually the same rate. Acetylene (0.1% v/v) instantaneously caused a complete inhibition of methane consumption, an observation that supports pMMO being the primary methane-oxidizing system. SolV could use both ammonium and nitrate as a nitrogen source. No growth occurred on ammonium without methane. Nitrogen fixation and anoxic nitrate-dependent methanol oxidation was not observed.

SolV has a typical K_s value for methane, namely $6 \mu\text{M}$. However, the affinity for oxygen was exceptionally high ($K_s 0.7 \mu\text{M}$), reflecting

Figure 2 | Growth characteristics of strain SolV. a, Typical growth curve showing decrease of methane (circles) and increase of optical density (triangles) at pH 2 and 55°C . The equation is the best exponential fit through the data points. **b**, Growth rate in relation to pH. The insert shows an enlargement of the data below pH 2. Different symbols indicate experiments performed on different days.

the need to compete for oxygen in its natural habitat, where microbial oxygen consumption and a constant flux of oxygen-depleted fumarolic gases, containing mainly carbon dioxide, will cause oxygen concentrations to be very low.

Fluorescence *in situ* hybridization (FISH) analysis of the isolate using the probe EUBIII, which is designed to mainly cover the Verrucomicrobiales^{24,25}, showed a strong hybridization signal (Supplementary Fig. 4). No signal was obtained with EUBI, EUBII or the alpha (ALF968), beta (BET42a) or gamma (GAM42a) proteobacterial probes^{25,26}. The Verrucomicrobia-like identity was confirmed by the sequence of its 16S rRNA gene obtained from pyrosequencing (see above). A specific probe was designed on the basis of this sequence (SolV830, see Methods) and used together with probe EUBIII to confirm the purity of the SolV culture. All cells from an exponentially growing culture showed double hybridization (Supplementary Fig. 4). Phylogenetic analysis of the 16S rRNA sequence of SolV indicated that the isolate represents the first member of a new subdivision within the Verrucomicrobia phylum (Fig. 3 and Supplementary Fig. 5). Pairwise distance analysis revealed <81% identity with members of other subdivisions^{6,27}.

Strain SolV is the first reported extreme acidiphilic methanotrophic bacterium and is phylogenetically placed outside the subphyla of the Alpha- and Gammaproteobacteria containing the established methanotrophs, and we propose to name it: '*Acidimethylosilex fumarolicum*', gen. nov. sp. nov. (Supplementary Information).

So far the Verrucomicrobia phylum contains only a few cultivated strains that are anaerobic or aerobic heterotrophs, growing on sugars in more or less complex media. However environmental clone libraries show that there is a large biodiversity of Verrucomicrobia and they are encountered in many ecosystems (soils, peat bogs, acid rock drainage and landfill leachate) often in relatively high numbers, but with an unknown physiology⁶. It is interesting to speculate that the widely distributed Verrucomicrobia phylum, from which most members remain uncultivated⁶, may be coupled to a geochemically relevant reaction. BLAST searches with the strain SolV 16S rRNA gene sequence showed very high identity (98–99%) to six environmental clones (Fig. 3) that were obtained during a geochemical study on microbial communities in acidic hot springs (Rainbow and Joseph's Coat) in Yellowstone National Park (unpublished; NCBI accession numbers: AY882698, AY882699, AY882710, AY882819, AY882820 and AY882834). This shows that bacteria similar to

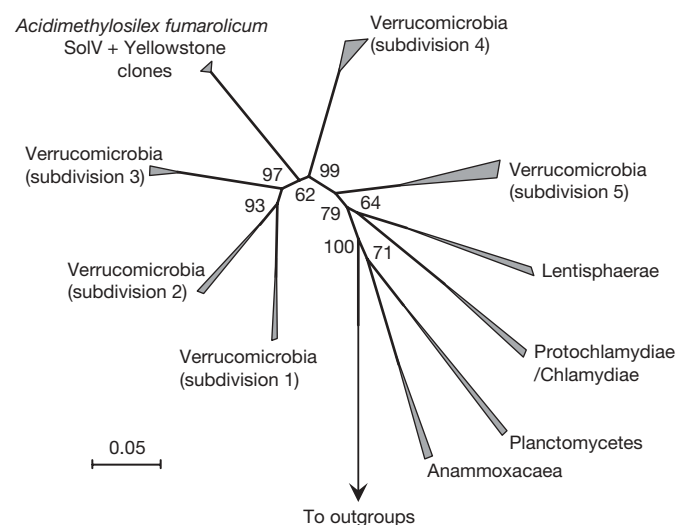


Figure 3 | Phylogenetic relationship between the 16S rRNA gene sequence of strain SolV and representatives of the Planctomycetes/Verrucomicrobia/Chlamydiae superphylum. The tree was calculated using the neighbour-joining algorithm with Kimura 2-parameter correction. Bootstrap values of 500 replicates are shown at the nodes. The scale bar represents 0.05 nucleotide changes per position.

A. fumarolicum may be common inhabitants of these extreme environments. The new *pmoA* and 16S rRNA gene sequences may help to identify the Planctomycetes/Verrucomicrobia/Chlamydiae superphylum methanotrophs from less extreme habitats and to show how they are globally distributed.

METHODS SUMMARY

Enrichments were started with mud and mixed soil samples from the Solfatara and incubated at 50 °C and pH 2.0 with methane as the sole source of energy and carbon. When methane consumption was observed, serial transfers into fresh medium were started. Finally a pure culture was obtained using the floating-filter technique¹⁷. Purity was checked by FISH and plating on medium enriched with yeast extract, without methane in the head space. DNA from environmental samples and genomic DNA from strain SolV was isolated as described²⁸. The *pmoA* and 16S rRNA genes were amplified by hot start using primers A189 and A682 (ref. 16), and 616F (5'-AGA GTT TGA TYM TGG CTC AG-3') and 630R (5'-CAK AAA GGA GGT GAT CC-3')²⁸, respectively. Pyrosequencing on genomic DNA was done as described¹⁸. FISH microscopy was performed as described²⁹ using the following nucleotide probes: EUBI, EUBII, EUBIII and SolV830 (5'-GGT CGA TTC CGC CAA CGC-3'). The latter probe was designed with the ARB program³⁰. Expression of *pmoA* mRNA was analysed by RT-PCR. The affinity for methane was estimated using cells from a batch culture ($OD_{600} = 0.24$). The apparent affinity constant for oxygen was estimated by measuring the methane respiration of a stirred culture in a 1 ml glass chamber equipped with a micro-oxygen sensor (Unisense A/S). Enzyme activities mentioned in Table 1 were measured according to ref. 23.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 July; accepted 4 September 2007.

Published online 14 November 2007.

- Dimitrov, L. Mud volcanoes — the most important pathway for degassing deeply buried sediments. *Earth Sci. Rev.* **59**, 49–76 (2002).
- Castaldi, S. & Tedesco, D. Methane production and consumption in an active volcanic environment of Southern Italy. *Chemosphere* **58**, 131–139 (2005).
- Etiopie, G. & Klusman, R. W. Geologic emissions of methane to the atmosphere. *Chemosphere* **49**, 777–789 (2002).
- Niemann, H. et al. Novel microbial communities of the Haakon Mosby mud volcano and their role as a methane sink. *Nature* **443**, 854–858 (2006).
- Alain, K. et al. Microbiological investigation of methane- and hydrocarbon-discharging mud volcanoes in the Carpathian Mountains, Romania. *Environ. Microbiol.* **8**, 574–590 (2006).
- Horn, M. & Wagner, M. The Planctomycetes, Verrucomicrobia, Chlamydiae and sister phyla comprise a superphylum with biotechnological and medical relevance. *Curr. Opin. Biotechnol.* **17**, 241–249 (2006).
- Houghton, J. Global warming. *Rep. Prog. Phys.* **68**, 1343–1403 (2005).
- Houghton, J. T., et al. *Climate Change 1995: The Science of Climate Change* (Cambridge Univ. Press, New York, 1995).
- Lösekann, T. et al. Diversity and abundance of aerobic and anaerobic methane oxidizers at the Haakon Mosby Mud Volcano, Barents Sea. *Appl. Environ. Microbiol.* **73**, 3348–3362 (2007).
- Hanson, R. S. & Hanson, T. E. Methanotrophic bacteria. *Microbiol. Rev.* **60**, 439–471 (1996).
- Dedysh, S. N. et al. *Methylocella tundreae* sp. nov., a novel methanotrophic bacterium from acidic tundra peatlands. *Int. J. Syst. Evol. Microbiol.* **54**, 151–156 (2004).
- Lieberman, R. L. & Rosenzweig, A. C. Crystal structure of a membrane-bound metalloenzyme that catalyses the biological oxidation of methane. *Nature* **434**, 177–182 (2005).
- Trotsenko, Y. A. & Khmelenina, V. N. Biology of extremophilic and extremotolerant methanotrophs. *Arch. Microbiol.* **177**, 123–131 (2002).
- Dedysh, S. N. et al. Isolation of acidophilic methane-oxidizing bacteria from northern peat wetlands. *Science* **282**, 281–284 (1998).
- Dedysh, S. N. et al. *Methylocystis heyeri* sp. nov., a novel type II methanotrophic bacterium possessing 'signature' fatty acids of type I methanotrophs. *Int. J. Syst. Evol. Microbiol.* **57**, 472–479 (2007).
- Holmes, A. J., Costello, A., Lidstrom, M. E. & Murrell, J. C. Evidence that participate methane monooxygenase and ammonia monooxygenase may be evolutionarily related. *FEMS Microbiol. Lett.* **132**, 203–208 (1995).
- Visser, J. M., Stefess, G. C., Robertson, L. A. & Kuenen, J. G. *Thiobacillus* sp. W5, the dominant autotrophic oxidizing sulfide to sulfur in a reactor for aerobic treatment of sulfidic wastes. *Antonie Van Leeuwenhoek* **72**, 127–134 (1997).
- Margulies, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
- Stoecker, K. et al. Cohn's *Crenothrix* is a filamentous methane oxidizer with an unusual methane monooxygenase. *Proc. Natl Acad. Sci. USA* **103**, 2363–2367 (2006).

20. Dunfield, P. F. *et al.* Isolation of a *Methylocystis* strain containing a novel *pmoA*-like gene. *FEMS Microbiol. Ecol.* **41**, 17–26 (2002).
21. Tchawa Yimga, M., Dunfield, P. F., Rieke, P., Heyer, J. & Liesack, W. Wide distribution of a novel *pmoA*-like gene copy among type II methanotrophs, and its expression in *Methylocystis* strain SC2. *Appl. Environ. Microbiol.* **69**, 5593–5602 (2003).
22. Ward, N. *et al.* Genomic insights into methanotrophy: The complete genome sequence of *Methylococcus capsulatus* (Bath). *PLoS Biol.* **2**, 1616–1619 (2004).
23. Dedysh, S. N. *et al.* *Methylocella palustris* gen. nov., sp. nov., a new methane-oxidizing acidophilic bacterium from peat bogs, representing a novel subtype of serine-pathway methanotrophs. *Int. J. Syst. Evol. Microbiol.* **50**, 955–969 (2000).
24. Daims, H., Bruhl, A., Amann, R., Schleifer, K.-H. & Wagner, M. The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Syst. Appl. Microbiol.* **22**, 434–444 (1999).
25. Loy, A., Maixner, F., Wagner, M. & Horn, M. probeBase — an online resource for rRNA-targeted oligonucleotide probes: new features 2007. *Nucleic Acids Res.* **35**, D800–D804 (2007).
26. Manz, W., Amann, R., Ludwig, W., Wagner, M. & Schleifer, K.-H. Phylogenetic oligodeoxy-nucleotide probes for the major subclasses of proteobacteria: Problems and solutions. *Syst. Appl. Microbiol.* **15**, 593–600 (1992).
27. Hugenholtz, P., Goebel, B. M. & Pace, N. R. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* **180**, 4765–4774 (1998).
28. Juretschko, S. *et al.* Combined molecular and conventional analyses of nitrifying bacterium diversity in activated sludge: *Nitrosococcus mobilis* and *Nitrospira*-like bacteria as dominant populations. *Appl. Environ. Microbiol.* **64**, 3042–3051 (1998).
29. Snaird, J., Amann, R., Huber, I., Ludwig, W. & Schleifer, K.-H. Phylogenetic analysis and *in situ* identification of bacteria in activated sludge. *Appl. Environ. Microbiol.* **63**, 2884–2896 (1997).
30. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (<http://www.arb-home.de>) (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank M. Strous and S. Castaldi for critical reading and discussion, L. van Niftrik and G.-J. Janssen for technical assistance with electron microscopy, M. Schmid for assistance with FISH microscopy and phylogenetic analyses, and H. A. Mohammadi and M. Gerrits for technical assistance in cultivation. H. Lunstroo is acknowledged for allowing access to the 454-sequencing technology, and G. Angarano for allowing access to the Solfatara and P. Mariani for assistance during sampling.

Author Contributions A.P. and D.T. performed the sampling; A.P. did the enrichment and isolation; K.H. and A.P. carried out the physiological experiments; K.H. and H.R.H. were responsible for the molecular analysis; A.P. and H.J.M.O.d.C. performed phylogenetic analyses, alignments and probe design. The research was conceived by A.P., M.S.M.J. and H.J.M.O.d.C. and was based on observations made by D.T. A.P., M.S.M.J., D.T. and H.J.M.O.d.C. contributed to interpreting the data and writing the paper.

Author Information The nucleotide sequence data have been deposited in GenBank under accession numbers EF591085 (*pmo_1*), EF591086 (*pmo_2*), EF591087 (*pmo_3*) and EF591088 (16S rRNA). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.S.M.J. (m.jetten@science.ru.nl) or H.J.M.O.d.C. (h.opdencamp@science.ru.nl).

METHODS

After 3 weeks, methane consumption was observed and repeated serial transfers of the mud and mixed-soil culture into fresh medium (see below) were started. Finally the culture was serially diluted and aseptically filtered through 25-mm polycarbonate filters (0.2 μm , Nucleopore), which were placed floating on medium in Petri dishes and incubated in closed jars under a methane atmosphere (see below)¹⁷. Tiny whitish colonies appeared on the filters after 1 week and microscopic observation revealed only one rod-shaped morphotype. Purity was checked by FISH and plating on medium enriched with yeast extract, without methane in the head space. No growth was observed on this medium.

Culture conditions. The culture medium was based on the Fangaia mineral concentrations and composed of 0.4 mM MgCl_2 , 2 mM CaHPO_4 , 1 mM Na_2SO_4 , 2 mM K_2SO_4 , 2 mM $(\text{NH}_4)_2\text{SO}_4$, 3% autoclaved liquid from the Fangaia mud pool at Pozzuoli; 1 ml l^{-1} trace elements (in mg l^{-1}) $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ (4.4), $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ (1.0), $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ (1.0), $(\text{NH}_4)_6\text{MO}_7\text{O}_{24} \cdot 4\text{H}_2\text{O}$ (0.22), $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ (0.32), $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$ (0.32). The pH was adjusted with H_2SO_4 or NaOH. Bacteria were grown in 120 ml serum bottles with 10 ml of medium and 2–5% CO_2 and 2–5% CH_4 in the headspace. Bottles were incubated at 50–55 °C on a rotary shaker at 250 r.p.m. To determine the reaction stoichiometry, gas samples were taken from triplicate cultures with a gaslock syringe and methane, carbon dioxide, oxygen and hydrogen were analysed on a HP 6890 gas chromatograph with a Porapak Q column and thermal conductivity detection. Yield on methane was determined by harvesting cells in the late exponential phase. Cells were centrifuged and washed with 1 mM HCl and dried under vacuum at 70 °C until constant weight.

***pmoA* and 16S rRNA gene sequence analysis.** DNA from environmental samples and genomic DNA from strain SolV was isolated as described²⁸ without the use of lytic enzymes. For *pmoA* PCR under non-restrictive conditions the annealing temperature was lowered from 56 °C to 48 °C. The products were purified from an agarose gel with the QIAEX II gel extraction kit (Qiagen) and cloned using the TOPO TA cloning kit (Invitrogen). Plasmids were purified with FlexiPrep kit (Amersham Biosciences) and sequenced with M13R and M13F primers, which flank the cloning site of the vector. Pyrosequencing on genomic DNA was done as described¹⁸.

Real-time RT-PCR analysis. Samples (50 ml at $OD_{600} = 0.85$) from methane-grown chemostat cultures were rapidly cooled and RNA was isolated using the Omega RNA extraction kit (Omega Bio-Tec) according to the manufacturer's protocol. Transcription products of *pmoA* were detected using SolV-specific primers (REVp1032 5'-GCAAARCTTCTCATYAGTWCC-5'; FORp1034 5'-GTGGATGAATCGGTATTGG-3'). Reverse transcription was performed with primer REVp1032 and RevertAid M_MuLV (Fermentas) Quantitative PCR was done using the iQ custom SYBR Green supermix kit (Bio-Rad), according to the manufacturer's instructions. The PCR program on the Biorad MyiQ was 3 min 95 °C and 40 cycles 30 s at 95 °C, 30 s at 54 °C, 30 s at 72 °C.

FISH microscopy. On the basis of the obtained 16S rRNA gene (see above) a new oligonucleotide probe (SolV830, 5'-GGT CGA TTC CGC CAA CGC-3') was designed using the probe-design software of the ARB program³⁰. Optimum formamide concentration for this probe was 20%.

Kinetics and enzyme activities. The affinity for methane was estimated by measuring the consumption rate in a series of incubations of 10 ml samples, taken from a batch culture ($OD_{600} = 0.24$) in 100 ml bottles at 50 °C. Various amounts of methane were added and bottles were shaken vigorously at 500 r.p.m. Virtually linear rates were measured during one hour. Rates were proportional to the cell density in the range used, indicating that there was no mass-transfer limitation for methane. Five ml of the culture were preincubated at 50 °C in a closed 100 ml bottle with 30 ml of methane to ensure excess methane compared to oxygen. The oxygen-consumption rates were calculated from the decrease in oxygen concentration over time.

Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia

Peter F. Dunfield^{1†}, Anton Yuryev², Pavel Senin^{3,4}, Angela V. Smirnova¹, Matthew B. Stott¹, Shaobin Hou^{3,4}, Binh Ly^{3,4}, Jimmy H. Saw³, Zheming Zhou⁵, Yan Ren⁵, Jianmei Wang⁵, Bruce W. Mountain¹, Michelle A. Crowe¹, Tina M. Weatherby⁶, Paul L. E. Bodelier⁷, Werner Liesack⁸, Lu Feng⁵, Lei Wang⁵ & Maqsoodul Alam^{3,4}

Aerobic methanotrophic bacteria consume methane as it diffuses away from methanogenic zones of soil and sediment¹. They act as a biofilter to reduce methane emissions to the atmosphere, and they are therefore targets in strategies to combat global climate change. No cultured methanotroph grows optimally below pH 5, but some environments with active methane cycles are very acidic^{2,3}. Here we describe an extremely acidophilic methanotroph that grows optimally at pH 2.0–2.5. Unlike the known methanotrophs, it does not belong to the phylum Proteobacteria but rather to the Verrucomicrobia, a widespread and diverse bacterial phylum that primarily comprises uncultivated species with unknown genotypes. Analysis of its draft genome detected genes encoding particulate methane monooxygenase that were homologous to genes found in methanotrophic proteobacteria. However, known genetic modules for methanol and formaldehyde oxidation were incomplete or missing, suggesting that the bacterium uses some novel methylotrophic pathways. Phylogenetic analysis of its three *pmoA* genes (encoding a subunit of particulate methane monooxygenase) placed them into a distinct cluster from proteobacterial homologues. This indicates an ancient divergence of Verrucomicrobia and Proteobacteria methanotrophs rather than a recent horizontal gene transfer of methanotrophic ability. The findings show that methanotrophy in the Bacteria is more taxonomically, ecologically and genetically diverse than previously thought, and that previous studies have failed to assess the full diversity of methanotrophs in acidic environments.

Methane is the second most important greenhouse gas, estimated to contribute 18% of the total atmospheric radiative forcing⁴. It is produced in anoxic environments primarily through the microbial degradation of organic matter¹, but abiogenic methane emitted from seeps and geothermal areas is also a major contribution to the atmospheric budget, estimated at 45–75 Tg annually^{2,5}. Hell's Gate (Tikitere), New Zealand, is a geothermal area rich in abiogenic methane⁶. We studied methane oxidation in an area of woody vegetation that had recently died as a result of the onset of steam emission from below. There was a temperature gradient in the soil, which consisted of an organic horizon (0–2 cm, pH 3.1, 31 °C), an A horizon (2–5 cm, pH 3.8, 42 °C), a B horizon (5–10 cm, pH 4.3, 53 °C), black ash (10–15 cm, pH 4.5, 63 °C) and pumice mixed with ash (below 15 cm, pH 4.7, 71 °C at 20 cm). Soil gas at a depth of 60 cm contained about 1% (v/v) methane. This did not decline linearly towards the soil surface. Instead it approached a minimum at 10–20 cm below the

surface, indicating that a subsurface sink consumed the methane as it diffused upwards (Fig. 1). Bacterial 16S rRNA genes amplified by polymerase chain reaction (PCR) from the soil layer at 10–15 cm depth were dominated (33 of 35 cloned genes) by a single sequence that showed low identity (less than 83%) to any cultivated bacterium and grouped phylogenetically within the phylum Verrucomicrobia (Supplementary Fig. 1). A bacterium possessing this 16S ribosomal RNA gene sequence (isolate V4) was subsequently isolated at 60 °C in air supplemented with 25% (v/v) methane as the sole energy source. Isolate V4 was able to grow and oxidize methane down to at least pH 1.5 (Fig. 2). The optimum was pH 2.0–2.5 (Fig. 3). The growth rate was near optimal between pH 1.5 and pH 3.0, but weak growth was observed as low as pH 1.0 and as high as pH 6.0.

Two properties of this bacterium are unique in comparison with all known methanotrophs: its phylogenetic placement in the phylum Verrucomicrobia and its extremely acidophilic phenotype.

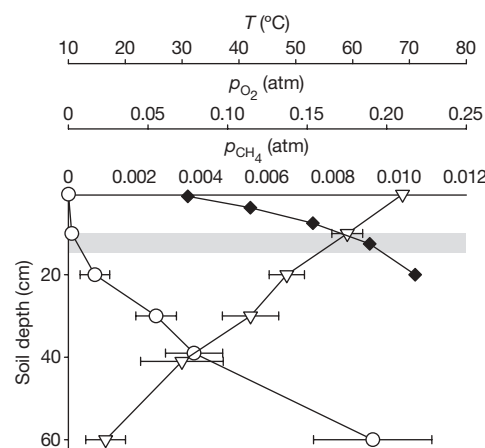


Figure 1 | Vertical profiles of methane partial pressures (open circles), O₂ partial pressures (open triangles) and temperature (filled diamonds) in a geothermal soil. Partial pressures are shown as means \pm s.e.m. ($n = 5$). Methane declined from 0.01 atm at 60 cm depth to below the detection limit (less than 100 p.p.m.) at 10–20 cm depth, suggesting that there was a subsurface methane sink. The linear decline in O₂ indicates that there was a deeper sink for this gas and that the diffusion rate was relatively constant in different soil layers. The grey area represents the soil sample (10–15 cm depth) from which bacterial 16S rRNA genes were amplified by PCR.

¹GNS Science, Extremophile Research Group, Private Bag 2000, Taupo, New Zealand. ²Ariadne Genomics, Inc., 9430 Key West Avenue no. 113, Rockville, Maryland 20850, USA. ³Department of Microbiology, University of Hawaii, Snyder Hall no. 207, 2538 The Mall, Honolulu, Hawaii 96822, USA. ⁴Advanced Studies in Genomics, Proteomics and Bioinformatics, College of Natural Sciences, Keller Hall 319, 2565 McCarthy Mall, Honolulu, Hawaii 96822, USA. ⁵TEDA School of Biological Sciences and Biotechnology, Nankai University, 23 HongDa Street, Tianjin 300457, China. ⁶Biological Electron Microscopy Facility, Pacific Biosciences Research Center, University of Hawaii at Manoa, Snyder Hall no. 118, Honolulu, Hawaii 96822, USA. ⁷Netherlands Institute of Ecology (NIOO-KNAW), Centre for Limnology, Department of Microbial Wetland Ecology, Rijksstraatweg 6, 3631 AC, Nieuwersluis, The Netherlands. ⁸Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Strasse, 35043 Marburg, Germany. [†]Present address: Department of Biological Sciences, University of Calgary, 2500 University Drive NW, Calgary, Alberta, T2N 1N4, Canada.

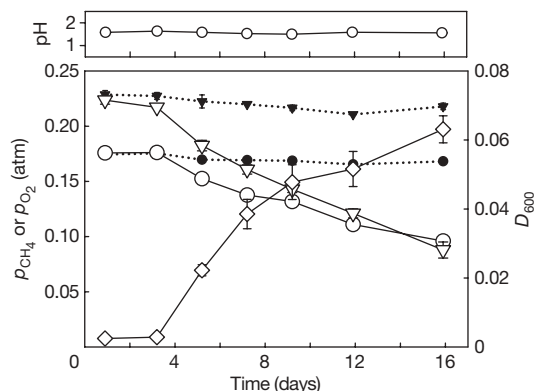


Figure 2 | D_{600} (open diamonds), methane consumption (open circles), and O_2 consumption (open triangles) of isolate V4 growing in liquid medium at pH 1.5 and 50 °C. Data are shown as means \pm s.e.m. for triplicate culture vials. Where error bars are not visible they are smaller than the symbols. The upper panel shows that the pH was constant throughout the experiment. The dotted lines represent partial pressures of methane (solid circles) and O_2 (solid triangles) in uninoculated blanks (means \pm s.e.m. for duplicate vials); the slight decline is due to removal during sampling.

Decades of previous study have identified only 13 genera of aerobic methanotrophs. All belong to the phylum Proteobacteria, in the classes Alphaproteobacteria and Gammaproteobacteria. They include two moderate acidophiles, *Methylocella* and *Methylocapsa*, which have pH optima of 5.0–5.5 and a lower limit of 4.0–4.5 (refs 3, 7). Isolate V4 is therefore by far the most acidophilic methanotroph yet found. Cells were non-motile rods with dimensions 0.3–0.5 $\mu\text{m} \times 1$ –4 μm . A tubular membrane inside an invagination of the inner membrane was observed in a small proportion of cells (Fig. 4). Intracytoplasmic membranes, usually occurring in parallel stacks, are common in methanotrophs and are suspected to house the membrane-bound particulate methane monooxygenase (pMMO) enzyme. The membrane structure in isolate V4 could serve the same function; however, its rare occurrence indicates that it may be only a stage in a morphological life cycle or that it may occupy a small volume of the cell.

Aerobic methanotrophic bacteria use monooxygenase enzymes to convert methane to methanol, which is then oxidized to formaldehyde, formate and CO_2 . Two forms of methane monooxygenase are known, a soluble form (sMMO) present in a few species, and a membrane-bound form (pMMO) present in all known genera except *Methylocella*⁸. Microbial ecologists have taken advantage of the nearly universal occurrence of pMMO to design detection assays specific to methanotrophs. Degenerate oligonucleotide primer sets have been employed to amplify *pmoA* genes (encoding a subunit of pMMO)

from various environments by means of PCR. On the basis of *pmoA* sequence phylogeny, individual species can then be identified^{9,10}. We were unable to amplify a *pmoA* gene product from a DNA extract of isolate V4, nor directly from a soil DNA extract, using standard primer sets and PCR protocols¹⁰. To obtain a better understanding of methane metabolism by this bacterium, we therefore sequenced and assembled a draft genome using an eightfold-coverage whole-shotgun approach¹¹.

Analysis of the draft genome (2.1 megabases) allowed the identification of genetic systems involved in methanotrophic growth. Three complete *pmoCAB* operons encoding pMMO were found, each with the same gene arrangement found in Proteobacteria^{12,13} and each with a putative upstream σ^{70} promoter sequence. Membrane topologies predicted from derived amino acid sequences were identical to the corresponding proteins in Proteobacteria¹³, although the maximum sequence identities were only 34–65% (Supplementary Table 1). The *pmoA* gene is a conserved and largely reliable phylogenetic marker^{9,10}, so a comparative phylogenetic analysis was made of the three derived PmoA sequences of isolate V4. These grouped into a single evolutionary lineage distinct from PmoA and AmoA sequences of other methane-oxidizing and ammonia-oxidizing microorganisms (Fig. 5), indicating an ancient divergence of Verrucomicrobia and Proteobacteria methanotrophs rather than a recent horizontal transfer of *pmo*.

Three divergent *pmo* operons have never before been detected in a methanotroph. *Methylococcus capsulatus* and some other species possess two nearly identical copies of *pmoCAB*^{12,13}, whereas two divergent operons have been found in a *Methylocystis* species¹⁴. Two of the derived PmoA sequences in isolate V4 were 90% identical (Supplementary Table 2), but copy 3 was less than 50% identical with the other two, suggesting that different selection pressures have acted on it. However, it is unlikely that any of the three operons encode an ammonia monooxygenase. The bacterium did not grow on ammonium, and an analysis of the three derived PmoA sequences for putative ‘signature’ amino acids found a higher proportion of methanotroph-specific than nitrifier-specific signatures in each¹⁵ (Supplementary Table 2). However, Hell’s Gate is rich in inorganic nitrogen⁶, so the three putative pMMO enzymes may be differentially sensitive to competitive inhibition by ammonia. Alternatively, they may have different kinetic properties or pH optima.

The genome was also searched for homologues of other methylo-trophy genes, especially those identified in the fully sequenced methanotroph *Methylococcus capsulatus*¹² (Supplementary Table 1). Besides pMMO, gene clusters encoding formate dehydrogenase and methylamine dehydrogenase were present, although several genes encoding accessory proteins for assembly and stabilization of the latter enzyme¹⁶ were not identified, and we were unable to

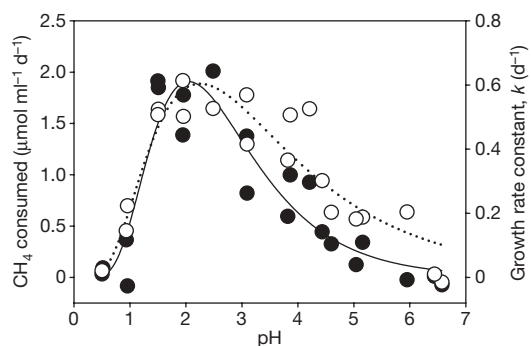


Figure 3 | Growth rate constant based on D_{600} (open circles) and average methane consumption rates (filled circles) of isolate V4 at a range of pH values. Each point represents the mean of two culture vials, measured over 7–14 days of incubation. The lines were estimated as log-normal iterative best fits to the data.

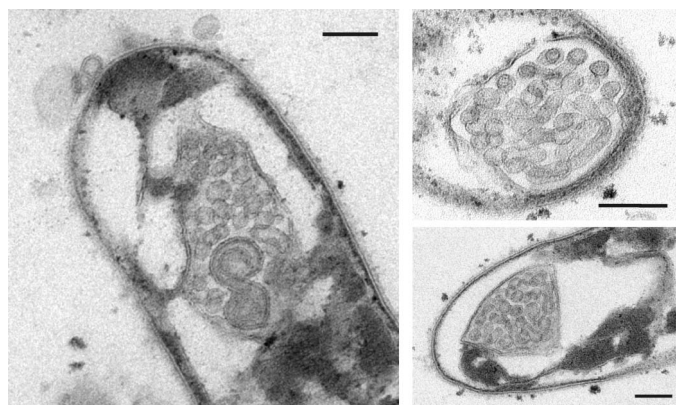


Figure 4 | Transmission electron micrographs of internal membrane structures observed in some cells of *Verrucomicrobia* isolate V4. A sac containing tubular membrane structures is attached to the inner cell membrane. Scale bar, 100 nm.

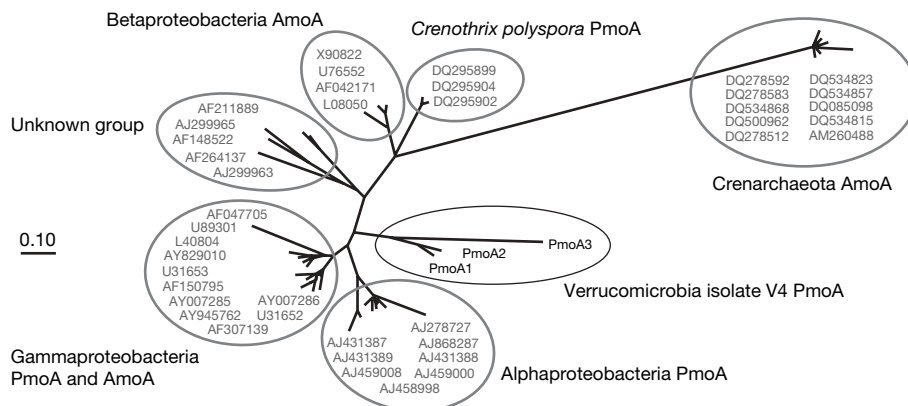


Figure 5 | Phylogenetic tree constructed from derived PmoA and AmoA sequences (subunits of particulate methane monooxygenase or ammonia monooxygenase), showing the relative positions of the three sequences from Verrucomicrobia isolate V4. Distinct groups corresponding to broad microbial taxa can be delineated, with the exception of Crenothrix polyspora,

which groups apart from other Gammaproteobacteria. The tree was constructed on the basis of 165 amino-acid positions by using TREE-PUZZLE²⁹, a quartet maximum-likelihood method. The support value from 10,000 puzzling steps for the branch to the Verrucomicrobia was 86%. The scale bar represents 0.1 change per amino-acid position.

cultivate the bacterium on methylamine. Other similarities to *Methylococcus* included genes encoding nitrite reductase, hydroxylamine oxidoreductase and nitric oxide reductase. These enzymes may remove toxic byproducts of competitive ammonia oxidation by pMMO.

However, there was no genomic evidence for some key enzymes of methylotrophy, including methanol dehydrogenase and tetrahydro-methanopterin enzymes for formaldehyde oxidation. Methanol is the immediate product of pMMO, so some form of methanol dehydrogenase should be present. However, no homologues were identified for the *mxhI* gene encoding the small subunit of this enzyme, nor for a variety of other essential genes¹⁶. Genomes of several methylotrophic and non-methylotrophic bacteria contain truncated *mxh*-like gene clusters similar to that in isolate V4, but it is doubtful whether these express a methanol dehydrogenase^{17,18}. To our knowledge, isolate V4 is also the first fully sequenced methylotroph to lack the tetrahydro-methanopterin pathway for formaldehyde oxidation. Enzymes of this pathway are taxonomically widespread, occurring in Alphaproteobacteria, Betaproteobacteria and Gammaproteobacteria^{12,16–18}, as well as in Archaea and Planctomycetes¹⁹. A potential alternative for formaldehyde oxidation is by means of tetrahydrofolate enzymes¹². Several are present in isolate V4 (Supplementary Table 1), but a complete pathway cannot be verified at this stage. Another alternative would be a glutathione-dependent or glutathione-independent formaldehyde dehydrogenase. Although we found no homologues to formaldehyde dehydrogenases from *Methylococcus capsulatus* or *Methylobacterium extorquens*, there were homologues to alcohol dehydrogenases that could conceivably oxidize formaldehyde.

Methanotrophs fix most of their carbon heterotrophically through serine or ribulose monophosphate pathways. No homologues were found for key enzymes (hexulose-6-phosphate synthase and hexulose-6-phosphate isomerase) of the ribulose monophosphate pathway^{20,21}. However, genes encoding some serine cycle enzymes were identified, indicating that a variant of this cycle may function (Supplementary Table 1). In addition, all genes necessary for a complete Calvin–Benson cycle were present. Carbon dioxide fixation contributes only a minor proportion of the total carbon assimilation of proteobacterial methanotrophs²⁰, but the genomic data imply an increased role in isolate V4. Experiments confirmed that the maximum growth rate (on methanol) was strongly dependent on CO₂. It decreased by two orders of magnitude when no CO₂ was added to culture vials, from 0.91 d^{−1} (s.e.m. 0.061) under 2–10% (v/v) CO₂ to less than 0.01 d^{−1} under ambient CO₂. A complete tricarboxylic acid cycle was also predicted, but we were unable to cultivate the bacterium on multicarbon substrates. Robust growth occurred only on the C₁ compounds methane and methanol. Although most

methanotrophs are obligately methylotrophic, the failure of isolate V4 to grow on short-chain organic acids such as acetate is curious, because these should be protonated under acidic growth conditions and diffuse freely across the cell membrane. Other factors must restrict growth on these substrates.

Some bacterial phyla, particularly the Verrucomicrobia and Acidobacteria, are widespread and abundant in nature but are represented by only a few isolates in culture collections²². Determining the metabolic lifestyles of these bacteria (for example, the recent discovery of photosynthesis in the Acidobacteria²³) is a major task facing microbial ecologists²². Stable-isotope-probing experiments using ¹³CH₄ to label bacterial DNA have identified Verrucomicrobia as potential members of methane-based food webs in an acidic soil²⁴ and an alkaline sediment²⁵, so methanotrophic members of this phylum may be ecologically diverse. We have recently obtained a methanotrophic mixed culture at 25 °C from an acidic mud (pH 3), which is dominated by a Verrucomicrobia showing less than 90% 16S rRNA gene sequence identity to isolate V4 (culture LP2A in Supplementary Fig. 1). Although the inoculum source was also a geothermal area, this finding does suggest that verrucomicrobial methanotrophs are a taxonomically broad group with varying environmental tolerances. Environmental surveys of methanotrophs based on *pmoA* recovery have failed to detect these methanotrophs in the past. Standard PCR primers^{9,10} show 3–13 mismatches with *pmoA* target regions in isolate V4, so this failure may be due to methodological limitations. Isolate V4 also possesses none of the membrane lipids considered unique to methanotrophs: 16:1ω8c, 16:1ω5t, or 18:1ω8c (ref. 26) (Supplementary Table 3), so studies targeting signature phospholipids have overlooked this group as well. The physiological and genomic analysis of our isolate, for which we propose the name '*Methylokorus infernorum*', will facilitate the design of more complete surveys of methanotrophs in different environments.

METHODS SUMMARY

Isolation. Soil crumbs were spread on plates of mineral salts medium at pH 4.5–5.5 and incubated at 60 °C in sealed jars containing 25% (v/v) CH₄ and 8% (v/v) CO₂ in air. Colony growth was restreaked onto new medium until pure. Purity was ensured by observation of colony and cell morphology and by genome sequencing of 43,008 reactions, which detected no contamination.

Characterization of isolate V4. Transmission electron microscopy²⁷ and phospholipid fatty acid analyses²⁶ were performed as described previously. For growth curve experiments, medium was adjusted to various pH values with H₂SO₄ and NaOH, and divided into aliquots in serum vials closed gas-tight with septa. CO₂ (2–5% v/v) and CH₄ (8–15%) were added. The pH, CH₄ and attenuation at 600 nm (*D*₆₀₀) were monitored at 2-day intervals. Methane was measured on a gas chromatograph equipped with a thermal conductivity detector (Varian Star 3600, 5 m × 0.75 mm Molecular Sieve 5A column, oven 35 °C, detector

270 °C). To test organotrophic growth, medium was supplemented with various sugars, organic acids, alcohols, methylamines or complex media at 0.05% (w/v). To test the effect of CO₂, cells were grown at pH 3.5 on methanol (2.5 ml l⁻¹), and 0, 2 or 10% (v/v) CO₂ was added to the headspace of vials.

Soil analyses. Soil gas was extracted into gas-tight syringes through a stainless steel tube. DNA was extracted with the PowerSoil Kit (Mo Bio Laboratories, Inc.). PCR amplification of 16S rRNA and *pmoA* genes, cloning with the TOPO TA Cloning Kit (Invitrogen) and DNA sequencing were performed as described previously^{10,11}.

Genome analyses. The genome was sequenced, assembled and mapped as described previously¹¹. Open reading frames were mapped into different metabolic pathways by using BLAST against the entire collection of bacterial protein sequences in the RefSeq database (National Centre for Biotechnology Information) and separately against the genome of *Methylococcus capsulatus* str. Bath. Orthologous pairs were calculated from BLAST results from the full-length sequence similarity, and best orthologues were estimated with a reciprocal best-hit method²⁸. The functional analysis was completed by manual annotation.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 August; accepted 24 October 2007.

Published online 14 November 2007.

- Conrad, R. Soil microorganisms as controllers of atmospheric trace gases (H₂, CO, CH₄, OCS, N₂O, and NO). *Microbiol. Rev.* **60**, 609–640 (1996).
- Castaldi, S. & Tedesco, D. Methane production and consumption in an active volcanic environment of southern Italy. *Chemosphere* **58**, 131–139 (2005).
- Dedysh, S. N. *et al.* Isolation of acidophilic methane-oxidizing bacteria from northern peat wetlands. *Science* **282**, 281–284 (1998).
- Intergovernmental Panel on Climate Change. *Climate Change 2007: The Physical Science Basis. Working Group I Contribution to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Solomon S. *et al.*) (Cambridge Univ. Press, Cambridge, 2007).
- Kvenvolden, K. A. & Rogers, B. W. Gaia's breath—global methane exhalations. *Mar. Petrol. Geol.* **22**, 579–590 (2005).
- Giggenbach, W. F., Sheppard, D. S., Robinson, B. W., Stewart, M. K. & Lyon, G. L. Geochemical structure and position of the Waiotapu geothermal field, New Zealand. *Geothermics* **23**, 599–644 (1994).
- Dedysh, S. N. Methanotrophic bacteria of acidic *Sphagnum* peat bogs. *Microbiology (Russia)* **71**, 638–650 (2002).
- Theisen, A. R. *et al.* Regulation of methane oxidation in the facultative methanotroph *Methylocella silvestris* BL2. *Mol. Microbiol.* **58**, 682–692 (2005).
- Dumont, M. G. & Murrell, J. C. Community-level analysis: key genes of aerobic methane oxidation. *Methods Enzymol.* **397**, 413–427 (2005).
- Knief, C., Lipski, A. & Dunfield, P. F. Diversity and activity of methanotrophic bacteria in different upland soils. *Appl. Environ. Microbiol.* **69**, 6703–6714 (2003).
- Hou, S. *et al.* Genome sequence of the deep-sea γ -proteobacterium *Idiomarina loihiensis* reveals amino acid fermentation as a source of carbon and energy. *Proc. Natl Acad. Sci. USA* **101**, 18036–18041 (2004).
- Ward, N. *et al.* Genomic insights into methanotrophy: the complete genome sequence of *Methylococcus capsulatus* (Bath). *PLoS Biol.* **2**, E303 (2004).
- Gilbert, B. *et al.* Molecular analysis of the *pmo* (particulate methane monooxygenase) operons from two type II methanotrophs. *Appl. Environ. Microbiol.* **66**, 966–975 (2000).
- Ricke, P., Erkel, C., Kube, M., Reinhardt, R. & Liesack, W. Comparative analysis of the conventional and novel *pmo* (particulate methane monooxygenase) operons from *Methylocystis* strain SC2. *Appl. Environ. Microbiol.* **70**, 3055–3063 (2004).
- Holmes, A. J. *et al.* Characterization of methanotrophic bacterial populations in soils showing atmospheric methane uptake. *Appl. Environ. Microbiol.* **65**, 3312–3318 (1999).
- Chistoserdova, L., Chen, S. W., Lapidus, A. & Lidstrom, M. L. Methylophily in *Methylobacterium extorquens* AM1 from a genomic point of view. *J. Bacteriol.* **185**, 2980–2987 (2003).
- Chistoserdova, L. *et al.* Genome of *Methylobacillus flagellatus*, molecular basis for obligate methylophily, and polyphyletic origin of methylophily. *J. Bacteriol.* **189**, 4020–4027 (2007).
- Kane, S. R. *et al.* Whole-genome analysis of the methyl tert-butyl ether-degrading beta-proteobacterium *Methylibium petroleiphilum* PM1. *J. Bacteriol.* **189**, 1931–1945 (2007).
- Chistoserdova, L. *et al.* The enigmatic *Planctomycetes* may hold a key to the origins of methanogenesis and methylophily. *Mol. Biol. Evol.* **21**, 1234–1241 (2004).
- Trotsenko, Y. Metabolic features of methane- and methanol-utilizing bacteria. *Acta Biotechnol.* **3**, 269–277 (2004).
- Kelly, D. P., Anthony, C. & Murrell, J. C. Insights from the complete genome sequence of the obligate methanotroph, *Methylococcus capsulatus*. *Trends Microbiol.* **13**, 195–198 (2005).
- Janssen, P. H. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl. Environ. Microbiol.* **72**, 1719–1728 (2006).
- Bryant, D. A. *et al.* *Candidatus Chloracidobacterium thermophilum*: an aerobic phototrophic *Acidobacterium*. *Science* **317**, 523–526 (2007).
- Morris, S. A., Radajewski, S., Willison, T. W. & Murrell, J. C. Identification of the functionally active methanotroph population in a peat soil microcosm by stable-isotope probing. *Appl. Environ. Microbiol.* **68**, 1446–1453 (2002).
- Lin, J. L. *et al.* Molecular diversity of methanotrophs in Transbaikian soda lake sediments and identification of potentially active populations by stable isotope probing. *Environ. Microbiol.* **6**, 1049–1060 (2004).
- Dedysh, S. N. *et al.* *Methylocystis heyerei* sp. nov., a novel type II methanotrophic bacterium possessing the 'signature' fatty acid of type I methanotrophs. *Int. J. Syst. Evol. Microbiol.* **57**, 472–479 (2007).
- Weatherby, T. M. & Lenz, P. H. Mechanoreceptors in calanoid copepods: Designed for high sensitivity. *Arthropod Struct. Dev.* **29**, 275–288 (2000).
- Ispolatov, I., Yuryev, A., Mazo, I. & Maslov, S. Binding properties and evolution of homodimers in protein–protein interaction networks. *Nucleic Acids Res.* **33**, 3629–3635 (2005).
- Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502–504 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Dedysh, J. C. Murrell and J. Euzéby for comments; A. Malahoff for supporting this research programme; and the Tikitere Trust for permission to sample at Hell's Gate. This work was supported in part by the Wairakei Environmental Mitigation Charitable Trust (P.D.). The genome analysis was funded by the US Department of Defense (M.A.).

Author Contributions P.F.D., B.W.M., M.A.C. and M.B.S. performed field sampling, methane measurement and molecular 16S rRNA analyses. P.F.D. and M.B.S. isolated and characterized the culture. S.H., B.L., J.H.S., Z.Z., Y.R., J.W., L.F., M.B.S., L.W., W.L. and M.A. conducted genome sequencing. P.F.D., P.S., A.Y., A.V.S., J.S., P.S. and M.A. conducted genome analyses. T.M.W. and M.B.S. performed electron microscopy. P.B. undertook phospholipid fatty-acid analysis.

Author Information Gene sequences referenced in this paper are deposited DDBJ/EMBL/GenBank under accession numbers AM900833–AM900834 and EU223838–EU223931. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.D. (p.dunfield@gns.cri.nz), M.A. (alam@hawaii.edu) or L.W. (wanglei@nankai.edu.cn).

METHODS

Isolation. The mineral salts medium V41 contained $0.1 \text{ g l}^{-1} \text{ NH}_4\text{Cl}$, $0.015 \text{ g l}^{-1} \text{ KH}_2\text{PO}_4$, $0.01 \text{ g l}^{-1} \text{ Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$, 10 mg l^{-1} yeast extract (added to the medium as a $0.2\text{-}\mu\text{m}$ -filter-sterilized solution after autoclaving), 3 ml l^{-1} FeEDTA solution (see below), 3 ml l^{-1} trace elements 1 solution and 1 ml l^{-1} trace elements 2 solution. The medium was adjusted to pH 5.5 and solidified by adding 22 g l^{-1} agar. The FeEDTA solution contained $1.54 \text{ g l}^{-1} \text{ FeSO}_4 \cdot 7\text{H}_2\text{O}$ and $2.06 \text{ g l}^{-1} \text{ Na}_2\text{EDTA}$. Trace elements solution 1 contained $0.44 \text{ g l}^{-1} \text{ ZnSO}_4 \cdot 7\text{H}_2\text{O}$, $0.20 \text{ g l}^{-1} \text{ CuSO}_4 \cdot 5\text{H}_2\text{O}$, $0.19 \text{ g l}^{-1} \text{ MnCl}_2 \cdot 4\text{H}_2\text{O}$, $0.06 \text{ g l}^{-1} \text{ Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, $0.10 \text{ g l}^{-1} \text{ H}_3\text{BO}_3$ and $0.08 \text{ g l}^{-1} \text{ CoCl}_2 \cdot 6\text{H}_2\text{O}$. Trace elements solution 2 contained 1.5 g l^{-1} nitrilotriacetic acid, $0.2 \text{ g l}^{-1} \text{ Fe}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$, $0.2 \text{ g l}^{-1} \text{ Na}_2\text{SeO}_3$, $0.1 \text{ g l}^{-1} \text{ CoCl}_2 \cdot 6\text{H}_2\text{O}$, $0.1 \text{ g l}^{-1} \text{ MnSO}_4 \cdot 2\text{H}_2\text{O}$, $0.1 \text{ g l}^{-1} \text{ Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, $0.1 \text{ g l}^{-1} \text{ Na}_2\text{WO}_4 \cdot 2\text{H}_2\text{O}$, $0.1 \text{ g l}^{-1} \text{ ZnSO}_4 \cdot 7\text{H}_2\text{O}$, $0.04 \text{ g l}^{-1} \text{ AlCl}_3 \cdot 6\text{H}_2\text{O}$, $0.025 \text{ g l}^{-1} \text{ NiCl}_2 \cdot 6\text{H}_2\text{O}$, $0.01 \text{ g l}^{-1} \text{ H}_3\text{BO}_3$ and $0.01 \text{ g l}^{-1} \text{ CuSO}_4 \cdot 5\text{H}_2\text{O}$ at pH 7.

Soil crumbs were spread on plates and these were placed in sealed jars containing 25% (v/v) CH_4 and 8% (v/v) CO_2 in air. Jars contained open vials of water to hydrate the air. Plates were incubated at 60°C and viewed at 2-week intervals for at least 12 weeks. Colonies that formed around soil crumbs were restreaked onto new medium. Optimization of the medium during this process indicated that pH 5.5 was too high for optimal growth, and the following altered medium composition (V42) was made: $0.4 \text{ g l}^{-1} \text{ NH}_4\text{Cl}$, $0.05 \text{ g l}^{-1} \text{ KH}_2\text{PO}_4$, $0.02 \text{ g l}^{-1} \text{ MgSO}_4 \cdot 7\text{H}_2\text{O}$ and $0.01 \text{ g l}^{-1} \text{ CaCl}_2 \cdot 6\text{H}_2\text{O}$, plus yeast extract and trace elements as above, adjusted to pH 4.5 and solidified with 15 g l^{-1} Phytigel plus $1 \text{ g l}^{-1} \text{ MgCl}_2 \cdot 6\text{H}_2\text{O}$. V42 medium was occasionally supplemented with 25 mg l^{-1} of a filter-sterilized vitamin mixture containing (per 100 mg): 0.8 mg folic acid, 8 mg vitamin B₁, 4 mg vitamin B₂, 1 mg niacin, 10 mg niacinamide, 15 mg pantothenate, 15 mg pyridoxine, 5 mg cobalamin, 5 mg biotin, 15 mg choline, 15 mg inositol and 7 mg *p*-aminobenzoic acid. The vitamin mixture resulted in no obvious stimulation of growth.

Colonies of the isolate V4 were light brown to copper in colour. Purity was ensured as described previously for methanotrophs³, by observation of colony growth on plates, by phase-contrast and transmission electron microscopy, by lack of growth on various heterotrophic substrates (see below), and by extraction of DNA, amplification and cloning of the 16S rRNA gene, and restriction-fragment-length polymorphism and sequencing analysis of 20 clones. All sequences were identical. Genome sequencing of 43,008 reactions detected no contaminating sequences (see below).

Transmission electron microscopy. Cells were fixed in 4% glutaraldehyde, postfixed in 1% osmium tetroxide, dehydrated in a graded ethanol series, and embedded in epoxy resin as described previously²⁷. Sections 80–90 nm thick were cut with a diamond knife on a Reichert Ultracut E ultramicrotome, collected on 200-mesh copper grids and stained with uranyl acetate and lead citrate. Sections were examined with a Zeiss/LEO 912 energy-filtering transmission electron microscope (EFTEM) operating at 100 kV, and images were collected with a Proscan 1k × 1k slow-scan frame-transfer charge-coupled device camera operated by Esivision software from Soft Imaging System. Images were saved as eight-bit uncompressed TIFF files and imported into Photoshop; histogram stretching was performed with the 'levels' setting, as permitted by the recommendations on digital manipulation set by the Microscope Society of America.

Growth experiments. For growth curve experiments, liquid medium V42 was adjusted to various pH values with H_2SO_4 and NaOH, and divided into serum vials closed gas-tight with rubber or Viton septa. CO_2 and CH_4 were added with syringes. Incubations were performed at 50°C . pH, CH_4 and D_{600} were monitored at 2-day intervals by the extraction of 2–3-ml culture samples with sterile syringes. Gas samples were removed by syringe for the measurement of methane on a gas chromatograph equipped with a thermal conductivity detector. The experiment shown in Fig. 2 used 500 ml of medium in 1-litre serum vials,

adjusted to an initial pH of 1.5. As there was a possibility of diffusion limitation of gaseous O_2 and CH_4 into the liquid phase when using large volumes as in the above experiment, growth rates over a range of pH values (Fig. 3) were tested using only 30-ml amounts of medium in 100-ml vials. Duplicate vials were adjusted to pH values of 0.5–7.0. One replicate was incubated statically; the other was shaken at 150 r.p.m. on a rotary shaker as a further control on potential diffusion limitation. Shaking resulted in only a small increase (less than 30%) in growth rate, so data from two replicates were combined into a single mean value. The growth rate constant (k^{-1}) was estimated from D_{600} values measured at 2-day intervals during the exponential growth phase (generally 1–7 days for rapid growth, 1–13 days for slower growth). Average methane consumption over this period was estimated by linear regression. The pH values given in Fig. 3 are average values measured over the incubation, in which the maximum change was observed between pH 4 and 5 (an average decrease of 0.5 pH unit per vial).

To test heterotrophic growth, liquid medium was supplemented with 0.05% (w/v) acetate, casamino acids, oxalic acid, citrate, malic acid, formate, benzoate, glucose, sucrose, ethanol, methanol, methylamine, trimethylamine, tryptophan, tryptic soy broth, propanol, nutrient broth, elemental sulphur, ascorbic acid, glycerol, xylose, arabinose, galactose, pectin, alginic acid or carboxymethylcellulose, or H_2 (10% v/v). When organic acids were added, the pH of the medium was readjusted to a control value (3.5 or 5.5 in different assays). All vials were supplemented with 5% (v/v) CO_2 . To test the effect of CO_2 on growth, cells were grown at pH 3.5 on methanol (2.5 ml l^{-1}), and 0, 2% or 10% (v/v) CO_2 was added to the headspace of vials.

The proposed name *Methylokorus infernorum* is derived from the Greek *methu* (wine) and *hulê* (wood), the Maori *korus* (a spiral) and the Latin *infernorum* (of infernal regions, because the strain was isolated from a location known as Hell's Gate).

Soil analyses. Gas samples were taken from soil by inserting to various depths a stainless steel tube with an outside diameter of 1/8 inch fitted with a Luer-Lok cap, and drawing out soil gas into syringes closed with valves. The gas volume in the tube was cleared before taking each sample. Extraction of DNA was performed with the PowerSoil Kit. Molecular PCR amplification of 16S rRNA and *pmoA* genes, cloning using the TOPO TA Cloning Kit and DNA sequencing were performed as described previously^{10,11}.

Genome analyses. Genome sequencing was performed as described previously¹¹. In brief, highly purified genomic DNA was extracted and used for 3-kilobase shotgun library construction. A total of 56 plates (384 plasmid preparations) were sequenced with forward and reverse primers (43,008 total sequencing reactions). Rough draft assembly was performed with the Arachne assembler. The sequence reads were trimmed and vector regions removed with the program Lucy. A custom-written Perl program was used to generate XML files needed to construct ancillary information needed for Arachne. Open reading frames were identified with Glimmer, and BLAST comparisons were made against the entire collection of bacterial protein sequences in the RefSeq database (National Centre for Biotechnology Information), and also separately against individual genomes of *Methylococcus capsulatus* str. Bath and *Nitrosococcus oceanus* ATCC 19707, and against methylophily genes in *Methylobacterium extorquens*¹⁶. The orthologous pairs were calculated from BLAST results by first calculating the full-length sequence similarity followed by calculation of best orthologues using a reciprocal best-hit method²⁸. The subsequent pathway analysis and protein functional annotation was performed by importing the genome into the bacterial database of Pathway Studio (Ariadne Genomics Inc.). This allowed the automatic transfer of protein annotation and pathway assignment from the annotated orthologues in other bacteria to V4 proteins. The functional analysis was completed with manual annotation and targeted BLAST searches. Membrane topologies were predicted for *pmoCAB* genes by using TMHMM (CBS), TMPred (EMBNet) and TopPred (Pasteur Institute).

The major *Vibrio cholerae* autoinducer and its role in virulence factor production

Douglas A. Higgins¹, Megan E. Pomianek², Christina M. Kraml³, Ronald K. Taylor⁴, Martin F. Semmelhack² & Bonnie L. Bassler^{1,5}

Vibrio cholerae, the causative agent of the human disease cholera, uses cell-to-cell communication to control pathogenicity and biofilm formation^{1,2}. This process, known as quorum sensing, relies on the secretion and detection of signalling molecules called autoinducers. At low cell density *V. cholerae* activates the expression of virulence factors and forms biofilms. At high cell density the accumulation of two quorum-sensing autoinducers represses these traits. These two autoinducers, cholerae autoinducer-1 (CAI-1) and autoinducer-2 (AI-2), function synergistically to control gene regulation, although CAI-1 is the stronger of the two signals. *V. cholerae* AI-2 is the furanosyl borate diester (2S,4S)-2-methyl-2,3,3,4-tetrahydroxytetrahydrofuran borate³. Here we describe the purification of CAI-1 and identify the molecule as (S)-3-hydroxytridecan-4-one, a new type of bacterial autoinducer. We provide a synthetic route to both the *R* and *S* isomers of CAI-1 as well as simple homologues, and we evaluate their relative activities. Synthetic (S)-3-hydroxytridecan-4-one functions as effectively as natural CAI-1 in repressing production of the canonical virulence factor TCP (toxin co-regulated pilus). These findings suggest that CAI-1 could be used as a therapy to prevent cholera infection and, furthermore, that strategies to manipulate bacterial quorum sensing hold promise in the clinical arena.

In a process called quorum sensing, bacterial communities track changes in their population densities by producing, releasing and detecting diffusible signalling molecules called autoinducers^{4,5}. Population-wide responses to the accumulation of autoinducers shape group behaviours such as the formation of biofilms, the expression of virulence factors, bioluminescence and sporulation^{6–9}. *V. cholerae* uses two parallel quorum-sensing systems to assess its population density (Fig. 1)¹. In one system the CqsS receptor responds to the molecule CAI-1, which is produced by the enzyme CqsA. In the second system, the LuxPQ receptor complex responds to the AI-2 molecule, made by the LuxS enzyme. CAI-1 is produced by several *Vibrio* species, which suggests that it functions as an intra-genus signal, whereas AI-2 is produced and detected by a wide variety of bacteria and is presumed to facilitate interspecies communication^{10,11}. AI-2 is a set of interconverting molecules all derived from the precursor (S)-4,5-dihydroxypentane-2,3-dione (DPD)^{12,13}. In the vibrios, the active AI-2 moiety is the furanosylborate diester (2S,4S)-2-methyl-2,3,3,4-tetrahydroxytetrahydrofuran borate (Fig. 1). CAI-1 has a much more marked influence on target gene expression than AI-2 and is therefore the major quorum-sensing signal in *V. cholerae*¹.

At low cell density, in the absence of autoinducers, *V. cholerae* expresses virulence factors and forms biofilms^{1,14,15}. This pattern of gene expression enables host colonization and contributes to persistence in the environment. In the presence of autoinducers, at high cell density, quorum sensing represses both the expression of virulence

factors and the formation of biofilms. These events are proposed to allow *V. cholerae* to leave the host, re-enter the environment in large numbers and initiate a new cycle of infection. Because autoinducers terminate rather than promote virulence in *V. cholerae*, activation of quorum sensing in *V. cholerae* by providing an autoinducer could form the basis of a strategy for therapeutic intervention in bacterial pathogenicity. To test the feasibility of manipulating quorum sensing for control of virulence, we identified, characterized and synthesized the major autoinducer, CAI-1, and tested whether we could use it to control quorum-sensing-regulated traits in *V. cholerae*.

To purify and identify CAI-1, we introduced the *cqsA* gene (encoding the CAI-1 synthase; see Fig. 1) into *Escherichia coli*, which is sufficient for the production and release of high-level CAI-1 activity into culture fluids^{1,10}. Recombinant *E. coli* produces substantially more CAI-1 activity than *V. cholerae* does in minimal medium, which simplified the initial purification and identification. CAI-1 activity was extracted from *E. coli* *pcqsA* cell-free culture fluids and purified

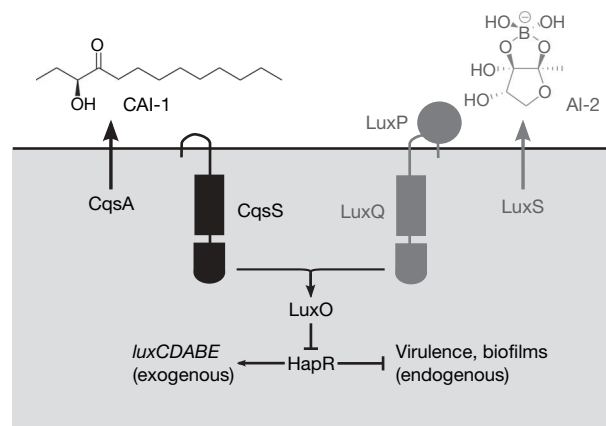


Figure 1 | Simplified model for quorum sensing in *V. cholerae*. The autoinducers CAI-1 ((S)-3-hydroxytridecan-4-one) and AI-2 ((2S,4S)-2-methyl-2,3,3,4-tetrahydroxytetrahydrofuran borate) are produced by the synthases CqsA and LuxS, respectively^{1,3}. CAI-1 is detected by the CqsS receptor; two proteins, LuxP and LuxQ, function together to detect AI-2. Information from both autoinducers is transduced through the LuxO protein to control the levels of the master transcription factor HapR. At low cell density, in the absence of autoinducers, HapR is not produced, so virulence factors are expressed and biofilms are formed. However, no light is produced from a luciferase (*luxCDABE*) cassette. At high cell density, in the presence of autoinducers, LuxO is inactivated, HapR is produced, and it represses genes for virulence factor production and biofilm formation while activating expression of the *luxCDABE* genes, resulting in bioluminescence. The complete circuit is reported in ref. 29.

¹Department of Molecular Biology, ²Department of Chemistry, ³Lotus Separations LLC, Princeton University, Princeton, New Jersey 08544, USA. ⁴Department of Microbiology and Immunology, Dartmouth Medical School, Hanover, New Hampshire 03755, USA. ⁵Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA.

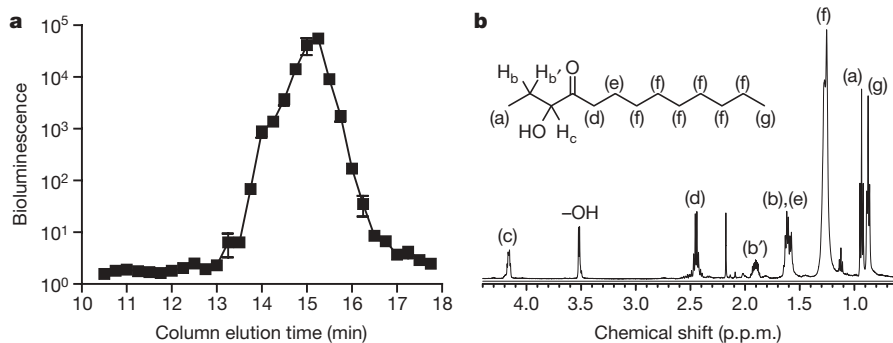


Figure 2 | Activity, ^1H -NMR spectrum and structural identification of CAI-1 as (S)-3-hydroxytridecan-4-one. **a**, CAI-1 was purified from cell-free extracts of *E. coli* *pqsA* by HPLC separation. Activity, as measured by bioluminescence production from a CAI-1-responsive *V. cholerae* reporter strain, was eluted as a single peak. Bioluminescence units are counts per second per 100 μl of culture. Error bars represent s.d. for triplicate samples.

by normal-phase high-performance liquid chromatography (HPLC). Activity was detected by using a *V. cholerae* reporter strain engineered to express bioluminescence exclusively in response to exogenously provided CAI-1 (ref. 1). CAI-1 activity appeared as one peak after separation by HPLC. Fractions in this peak contained 10,000-fold higher CAI-1 activity than control samples (Fig. 2a).

The structure of the dominant component (more than 90%) was assigned as the α -hydroxyketone shown in Fig. 2b, on the basis of spectroscopic evidence. A D_2O -exchangeable proton at $\delta = 3.5$ p.p.m. and an infrared absorbance at $3,490\text{ cm}^{-1}$ indicated the presence of a hydroxyl group. Using ^1H double-quantum NMR spectroscopy, a correlation between the hydroxyl group and proton (c), together with correlations between protons (bb') and protons (a) and (c), respectively, established a $\text{CH}_3\text{CH}_2\text{CH}(\text{OH})$ - unit. In addition, the correlation of protons (d) and (e), and of (e) and (g), with the amorphous signal for (f) indicated a long unbranched alkyl chain of uncertain length. The carbonyl group connecting these two fragments was confirmed by ^{13}C -NMR spectroscopy (212 p.p.m.) and IR ($1,709\text{ cm}^{-1}$). GC-MS analysis showed three components (Supplementary Fig. 1). The main component was assigned to the compound in which (f) corresponds to six methylene units (12 H) using both molecular ion and fragmentation information. From these data we concluded that the predominant molecule in the active fraction was $\text{CH}_3\text{CH}_2\text{CH}(\text{OH})\text{CO}(\text{CH}_2)_8\text{CH}_3$ (3-hydroxytridecan-4-one; chirality undefined). GC-MS analysis indicated that the active extract also contained small amounts of molecules in which the C_{13} acyl chain of the 13-carbon main component was replaced by a C_9 acyl or a C_8 acyl unit (hereafter represented by C_{13} , C_{12} and C_{11}).

To establish the structure of CAI-1, we synthesized and purified (more than 99% purity) each of the six candidates from the above analysis (that is, *R* and *S* $\text{CH}_3\text{CH}_2\text{CH}(\text{OH})\text{CO}(\text{CH}_2)_n\text{CH}_3$, where $n = 6, 7$ and 8 ; see Supplementary Scheme 1). The enantiomeric purity of each molecule was established by chiral chromatography. We verified that the components in the *E. coli* extract were the C_{13} , C_{12} and C_{11} homologues by comparison of GC-MS data with synthetic standards (Supplementary Figs 1 and 2). Furthermore, using chiral chromatographic methods developed with our synthetic standards, we confirmed that the dominant molecule made by *E. coli* was the *S* stereoisomer of 3-hydroxytridecan-4-one (Supplementary Fig. 4). Activity assays of synthetic CAI-1 and homologues support the identification of (S)-3-hydroxytridecan-4-one as CAI-1 (Fig. 3). All six of the synthetic CAI-1 homologues activated the *V. cholerae* quorum-sensing circuit to different degrees, following the pattern $\text{C}_{13}(\text{S}) > \text{C}_{13}(\text{R}) > \text{C}_{12}(\text{S}) \approx \text{C}_{12}(\text{R}) > \text{C}_{11}(\text{S}) > \text{C}_{11}(\text{R})$. This ranking shows that the molecules with longer acyl chains are more active than those with shorter acyl chains, and that, in general, a compound with *S* stereochemistry at the C3 position has greater activity than the

counterpart with *R* stereochemistry. Finally, by capitalizing on the purification and analysis strategies that we developed with our synthetic molecules and the *E. coli* extracts, we were able to obtain sufficient activity from *V. cholerae* for analysis. GC-MS and chiral chromatography showed that the only species present in *V. cholerae* culture fluids was the C_{13} version of the *S* isomer, establishing (S)-3-hydroxytridecan-4-one as the genuine CAI-1 molecule (Supplementary Figs 3 and 4).

A principal goal of quorum-sensing research is to develop strategies for the artificial manipulation of quorum-sensing-controlled processes in bacteria. Towards this end, we treated *V. cholerae* with synthetic CAI-1 and examined the consequences on production of TcpA. TcpA, a subunit of the toxin co-regulated pilus (TCP), is a primary host colonization factor¹⁶. Consistent with earlier results showing that virulence factors are expressed at low cell density and repressed by autoinducers at high cell density, high-level production of TCP occurred in *V. cholerae* mutants 'locked' in low-cell-density mode (ΔhapR), whereas cells 'locked' in high-cell-density mode (ΔluxO) produced no TcpA (Fig. 4a, and see Fig. 1)^{1,2}. The addition of synthetic CAI-1 at concentrations up to $10\text{ }\mu\text{M}$ substantially repressed TcpA production in a ΔluxQ , ΔcqsA *V. cholerae* strain that cannot respond to AI-2 or produce CAI-1 (Fig. 4b, top row). TcpA levels also decreased, although to a smaller extent, when CAI-1 was

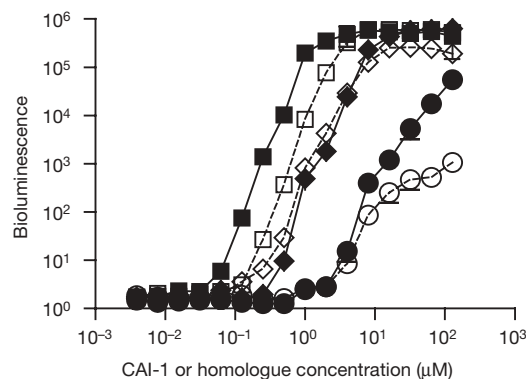


Figure 3 | Activity profiles for synthetic CAI-1 and homologues. *S* and *R* isomers of synthetic CAI-1 and shorter-length homologues identified in the *E. coli* extract were synthesized and tested for the ability to stimulate bioluminescence production (counts per second per 100 μl of culture) in the *V. cholerae* CAI-1 reporter strain. Symbols are as follows: (S)- and (R)-3-hydroxytridecan-4-one (C_{13}), filled and open squares, respectively; (S)- and (R)-3-hydroxydodecan-4-one (C_{12}), filled and open diamonds, respectively; (S)- and (R)-3-hydroxyundecan-4-one (C_{11}), filled and open circles, respectively. Error bars represent s.d. for triplicate samples.

provided to the $\Delta luxQ$, $cqsA^+$ strain (Fig. 4b, middle row). Taken together, these results show that exogenously supplied synthetic CAI-1 is capable of controlling gene expression in *V. cholerae*, both in the absence and in the presence of endogenously produced CAI-1. Synthetic CAI-1 had no effect on TcpA production in a $\Delta luxQ$, $\Delta cqsS$ strain, which cannot respond to either CAI-1 or AI-2 (Fig. 4b, bottom row). This final result demonstrates that information flow through the *V. cholerae* quorum-sensing circuit is absolutely required for the repression of TCP production by CAI-1, and, we infer, for the repression of other virulence factors.

Our identification of CAI-1 shows that it is an α -hydroxyketone and is therefore a new type of autoinducer. The standard activity curve we developed with synthetic CAI-1 allows us to estimate the autoinducer concentration in cell-free fluids derived from high-cell-density *V. cholerae* cultures to be 1.25 μM (data not shown). This is well in line with the concentrations and signalling activities of other bacterial autoinducers, which range from low micromolar (homoserine lactones) to nanomolar (AI-2)^{8,17}. The biological activity of CAI-1 is sensitive to side-chain length, as the 13-carbon molecule has eightfold greater activity than the 12-carbon molecule, which similarly is eightfold more active than the 11-carbon molecule (Fig. 3). The configuration of the hydroxyl group is also important, but to a smaller extent (roughly twofold differences). The relatively small effect of the configuration of the C3 hydroxyl group on activity is perhaps surprising, given that there are only two functional groups (hydroxyl and carbonyl) on the molecule. We are currently exploring the limits of, and requirements for, CAI-1 recognition by the CqsS receptor through the synthesis of related molecules, with a particular focus on functional groups and stereochemistry at the C3 and C4 positions as well as backbone length.

CqsA expression in *E. coli* leads predominantly to production of (S)-3-hydroxytridecan-4-one (13 carbons), with trace amounts of 3-hydroxydodecan-4-one (12 carbons) and 3-hydroxyundecan-4-one (11 carbons). *V. cholerae*, however, produces the 13-carbon S stereoisomer exclusively. Our findings suggest that CqsA couples a standard ten-carbon fatty-acyl unit to a three-carbon moiety to produce CAI-1 (discussed further below). We suspect that high-level

production of CqsA in heterologous *E. coli* exhausts the supply of the ten-carbon acyl substrate, enabling small amounts of shorter fatty-acyl chains (for example C₈) to be incorporated into the molecule.

The biochemical mechanism underlying the production of CAI-1 by CqsA is currently unknown. Database analysis shows that CqsA is similar to pyridoxal phosphate (PLP)-binding aminotransferases, with which AONS (8-amino-7-oxononanoate synthetase) shares the greatest homology. Related PLP-dependent aminotransferases typically catalyse condensations between amino acids and carboxylic thioesters, but the repertoire of PLP-dependent enzymes is diverse¹⁸. None, however, is known to generate an α -hydroxyketone such as CAI-1 or an α -diketone that could be a precursor to CAI-1. Our finding that introduction of *cqsA* into *E. coli* is sufficient for autoinducer production suggests either that CAI-1 is made from commonly occurring metabolic intermediates and only one enzyme (namely CqsA) is required or, alternatively, that *E. coli* provides the additional enzyme(s) needed to synthesize the molecule. If the former is correct, CqsA would be the first PLP-dependent aminotransferase to perform a generic coupling of carbon units to produce an α -hydroxyketone. *In vitro* studies are currently underway to define the biosynthetic mechanism of CAI-1 production.

The two largest families of autoinducer signals, the homoserine lactones used by Gram-negative bacteria and the oligopeptides used by Gram-positive bacteria, contain numerous related molecules, each possessing significant species specificity¹⁹. By contrast, AI-2 functions generically across species by the spontaneous derivitization of a shared precursor, DPD^{5,11}. Finally, there are particular autoinducer signals (and their corresponding synthases) that have unique structures and are produced by, and active only in, closely related organisms, such as the quinolones (pseudomonads), the γ -butyrolactones (streptomycetes), 3-hydroxyalmitic acid methyl ester (*Ralstonia solanacearum*), the diffusible signal factors (DSF; Xanthomonadaceae) and now the α -hydroxyketone CAI-1 (vibrios) (see Supplementary Fig. 5a for structures)^{20–24}. As a result of their limited distribution, clues to their existence are not revealed by comparison of diverse genome sequences. Rather, each of these molecules was individually discovered on the basis of a particular phenotype in a particular bacterium. We therefore suspect the existence of additional potent quorum-sensing signals of unique structure and restricted ranges of target organisms.

The molecule (S)-3-hydroxytridecan-4-one was previously unknown in biology. However, other α -hydroxyketones are known to have key functions in biological signalling systems. For example, the male coffee white stemborer *Xylotrechus quadripes* Chevrolat uses the α -hydroxyketone (S)-2-hydroxydecan-3-one as a sex pheromone²⁵. Interestingly, females are attracted specifically to the S isomer of this compound. Similar α -hydroxyketones are produced and used as pheromones by other insects (see Supplementary Fig. 5b)^{26,27}. To our knowledge, the biosynthetic steps responsible for production of these insect pheromones have not been defined and therefore cannot be compared with steps in CAI-1 synthesis. However, the similarities between the molecules raise the possibility that their syntheses as well as their functions in communication arose from a distant common ancestor.

TCP immunoblots show that CAI-1 is capable of blocking the expression of virulence factors, and *V. cholerae* mutants 'locked' at high cell density are completely avirulent^{1,2}. This is noteworthy, given the prevalence of cholera in the developing world and the rise of antibiotic-resistant strains of *V. cholerae*²⁸. That the autoinducer itself, and not an autoinducer antagonist, is an inhibitor of pathogenicity factors is a peculiarity of the *V. cholerae* disease process. Most pathogenic bacteria mount a long-lasting persistent infection. In cases in which quorum sensing is involved, the accumulation of autoinducers at high cell density promotes the expression of virulence factors⁷. *V. cholerae*, by contrast, elicits a self-limiting disease. At high cell density the bacteria are flushed from the host back into

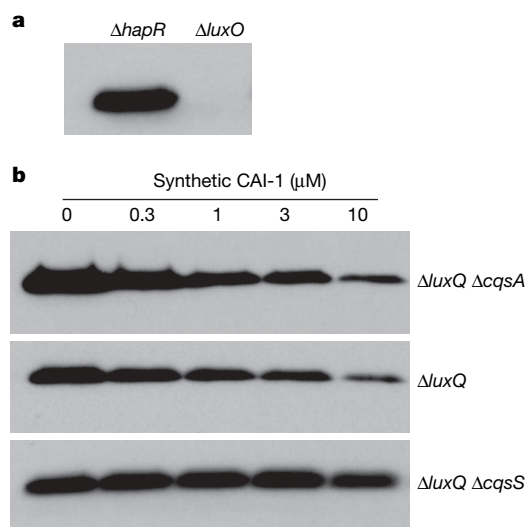


Figure 4 | CAI-1 inhibits virulence factor expression. Western blots were used to detect TcpA. **a**, TcpA is produced in the $\Delta hapR$ strain, which is 'locked' in low-cell-density mode, but not in the $\Delta luxO$ strain, which is 'locked' in high-cell-density mode. **b**, Increasing concentrations of synthetic CAI-1 repress TcpA production in a *V. cholerae* $\Delta luxQ$, $\Delta cqsA$ strain (which is incapable of responding to AI-2 and incapable of producing CAI-1) (top panel) and a *V. cholerae* $\Delta luxQ$ strain (which is incapable of responding to AI-2 but capable of producing CAI-1) (middle panel) but not in a *V. cholerae* $\Delta luxQ$, $\Delta cqsS$ strain (which is incapable of responding to both AI-2 and CAI-1) (bottom panel).

the environment. Consistent with this is the fact that autoinducer accumulation terminates the expression of virulence factors. This reciprocal relationship between autoinducer concentration and virulence in *V. cholerae* provides a unique opportunity to exploit quorum sensing to control pathogenicity. Moreover, the simplicity and inherent stability of CAI-1 may prove to be principal virtues in the development of future applications for bacterial control (see Supplementary Fig. 6). We argue that this work provides a demonstration that interference with quorum-sensing processes in general (either agonistic or antagonistic) has great promise in the clinical setting.

METHODS SUMMARY

CAI-1 purification. Autoinducer was purified from cell-free culture fluids of *E. coli* DH5 α harbouring inducible *V. cholerae* *cqsA* (VCA0523) on pTrc99A or from *V. cholerae* KSK1052 (El Tor C6706str $\Delta luxS$) as specified. Expression of *cqsA* in *E. coli* was induced with isopropyl β -D-thiogalactoside. After centrifugation and filtration, cell-free culture fluid was extracted with dichloromethane, evaporated, and assayed for CAI-1 activity with the use of the bioluminescence assay described below. After crude purification on silica gel, a concentrated extract (in dichloromethane) was injected onto a 2 cm \times 25 cm ethyl-pyridine HPLC column and eluted with an increasing concentration gradient of ethyl acetate in hexanes. Fractions were assayed for CAI-1 activity as described below. **Chemical and analytical methods.** ^1H -NMR spectra were recorded with a Varian Unity spectrometer (400 MHz), a Varian Unity/INOVA spectrometer (500 MHz) or a Bruker Avance II spectrometer (500 MHz). Chemical shifts were calibrated against a chloroform internal standard. Full synthesis methodology as well as details of ^1H -NMR, ^{13}C -NMR, GC-MS and supercritical fluid chromatography analyses are included in Supplementary Methods.

Bioluminescence assays. When assaying column fractions for autoinducer activity, 5% of each fraction was evaporated and dissolved in a water/acetonitrile mixture (2:1). This preparation was added at 2% volume to the CAI-1 reporter strain MM920 (*V. cholerae* El Tor C6706str $\Delta cqsA \Delta luxQ$ pBB1 (*luxCDABE* from *V. harveyi*)) as described previously¹. Synthetic CAI-1 preparations were dissolved in dimethylsulphoxide and supplied at 0.5% total volume to MM920 at specified concentrations.

Western blot analysis. *V. cholerae* strains were grown in AKI medium and samples were prepared as described previously¹. Anti-TcpA antibody and anti-rabbit IgG horseradish peroxidase conjugate (Promega) were used with the SuperSignal West Pico Chemiluminescent Substrate (Pierce) system.

Received 12 August; accepted 19 September 2007.

Published online 14 November 2007.

1. Miller, M. B., Skorupski, K., Lenz, D. H., Taylor, R. K. & Bassler, B. L. Parallel quorum sensing systems converge to regulate virulence in *Vibrio cholerae*. *Cell* **110**, 303–314 (2002).
2. Zhu, J. *et al.* Quorum-sensing regulators control virulence gene expression in *Vibrio cholerae*. *Proc. Natl Acad. Sci. USA* **99**, 3129–3134 (2002).
3. Chen, X. *et al.* Structural identification of a bacterial quorum-sensing signal containing boron. *Nature* **415**, 545–549 (2002).
4. Miller, M. B. & Bassler, B. L. Quorum sensing in bacteria. *Annu. Rev. Microbiol.* **55**, 165–199 (2001).
5. Waters, C. M. & Bassler, B. L. Quorum sensing: cell-to-cell communication in bacteria. *Annu. Rev. Cell Dev. Biol.* **21**, 319–346 (2005).
6. Davies, D. G. *et al.* The involvement of cell-to-cell signals in the development of a bacterial biofilm. *Science* **280**, 295–298 (1998).
7. Passador, L., Cook, J. M., Gambello, M. J., Rust, L. & Igilewski, B. H. Expression of *Pseudomonas aeruginosa* virulence genes requires cell-to-cell communication. *Science* **260**, 1127–1130 (1993).
8. Eberhard, A. *et al.* Structural identification of autoinducer of *Photobacterium fischeri* luciferase. *Biochemistry* **20**, 2444–2449 (1981).
9. Solomon, J. M., Lazazzera, B. A. & Grossman, A. D. Purification and characterization of an extracellular peptide factor that affects two different developmental pathways in *Bacillus subtilis*. *Genes Dev.* **10**, 2014–2024 (1996).

10. Henke, J. M. & Bassler, B. L. Three parallel quorum-sensing systems regulate gene expression in *Vibrio harveyi*. *J. Bacteriol.* **186**, 6902–6914 (2004).
11. Xavier, K. B. & Bassler, B. L. LuxS quorum sensing: more than just a numbers game. *Curr. Opin. Microbiol.* **6**, 191–197 (2003).
12. Schauder, S., Shokat, K., Surette, M. G. & Bassler, B. L. The LuxS family of bacterial autoinducers: biosynthesis of a novel quorum-sensing signal molecule. *Mol. Microbiol.* **41**, 463–476 (2001).
13. Miller, S. T. *et al.* *Salmonella typhimurium* recognizes a chemically distinct form of the bacterial quorum-sensing signal AI-2. *Mol. Cell* **15**, 677–687 (2004).
14. Hammer, B. K. & Bassler, B. L. Quorum sensing controls biofilm formation in *Vibrio cholerae*. *Mol. Microbiol.* **50**, 101–104 (2003).
15. Zhu, J. & Mekalanos, J. J. Quorum sensing-dependent biofilms enhance colonization in *Vibrio cholerae*. *Dev. Cell* **5**, 647–656 (2003).
16. Taylor, R. K., Miller, V. L., Furlong, D. B. & Mekalanos, J. J. Use of *phoA* gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. *Proc. Natl Acad. Sci. USA* **84**, 2833–2837 (1987).
17. Neiditch, M. B. *et al.* Ligand-induced asymmetry in histidine sensor kinase complex regulates quorum sensing. *Cell* **126**, 1095–1108 (2006).
18. Eliot, A. C. & Kirsch, J. F. Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu. Rev. Biochem.* **73**, 383–415 (2004).
19. Taga, M. E. & Bassler, B. L. Chemical communication among bacteria. *Proc. Natl Acad. Sci. USA* **100**, 14549–14554 (2003).
20. Pesci, E. C. *et al.* Quinolone signaling in the cell-to-cell communication system of *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA* **96**, 11229–11234 (1999).
21. Horinouchi, S. & Beppu, T. A-factor as a microbial hormone that controls cellular differentiation and secondary metabolism in *Streptomyces griseus*. *Mol. Microbiol.* **12**, 859–864 (1994).
22. Flavier, A. B., Clough, S. J., Schell, M. A. & Denny, T. P. Identification of 3-hydroxypalmitic acid methyl ester as a novel autoregulator controlling virulence in *Ralstonia solanacearum*. *Mol. Microbiol.* **26**, 251–259 (1997).
23. Wang, L. H. *et al.* A bacterial cell-cell communication signal with cross-kingdom structural analogues. *Mol. Microbiol.* **51**, 903–912 (2004).
24. Colnaghi Simionato, A. V., da Silva, D. S., Lambais, M. R. & Carrilho, E. Characterization of a putative *Xylella fastidiosa* diffusible signal factor by HRGC-El-MS. *J. Mass Spectrom.* **42**, 490–496 (2007).
25. Hall, D. R. *et al.* Identification of components of male-produced pheromone of coffee white stem borer, *Xylotrechus quadripes*. *J. Chem. Ecol.* **32**, 195–219 (2006).
26. Sakai, T., Nakagawa, Y., Takahashi, J., Iwabuchi, K. & Ishii, K. Isolation and identification of the male sex pheromone of the grape borer *Xylotrechus pyrrhoderus* Bates (Coleoptera: Cerambycidae). *Chem. Lett. (Jpn)* **13**, 263–264 (1984).
27. Leal, W. S., Shi, X., Nakamura, K., Ono, M. & Meinwald, J. Structure, stereochemistry, and thermal isomerization of the male sex pheromone of the longhorn beetle *Anaglyptus subfasciatus*. *Proc. Natl Acad. Sci. USA* **92**, 1038–1042 (1995).
28. Sack, D. A., Sack, R. B., Nair, G. B. & Siddique, A. K. Cholera. *Lancet* **363**, 223–233 (2004).
29. Hammer, B. K. & Bassler, B. L. Inaugural article: Regulatory small RNAs circumvent the conventional quorum sensing pathway in pandemic *Vibrio cholerae*. *Proc. Natl Acad. Sci. USA* **104**, 11145–11149 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Bassler, Semmelhack and Hughson groups for discussions. We thank I. Pelczar for the double-quantum filtered correlation spectroscopy experiment and initial ^{13}C -NMR studies. This work was supported by the Howard Hughes Medical Institute (B.L.B.) and grants from the National Science Foundation (B.L.B.) and the National Institutes of Health (B.L.B. and M.F.S.). M.E.P. was supported by a graduate fellowship from Amgen through the Medicinal Chemistry Division of the American Chemical Society and by the Horst Witzel Prize from the Cephalon Corporation.

Author Contributions D.A.H. purified CAI-1 and performed biological activity and virulence assays. M.E.P. performed chemical analyses and prepared synthetic CAI-1. C.M.K. aided in CAI-1 purification and conducted chiral chromatographic analyses.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to B.L.B. (bbassler@princeton.edu).

Localization of type 1 diabetes susceptibility to the MHC class I genes *HLA-B* and *HLA-A*

Sergey Nejentsev^{1*}, Joanna M. M. Howson^{1*}, Neil M. Walker¹, Jeffrey Szeszko¹, Sarah F. Field¹, Helen E. Stevens¹, Pamela Reynolds¹, Matthew Hardy¹, Erna King¹, Jennifer Masters¹, John Hulme¹, Lisa M. Maier¹, Deborah Smyth¹, Rebecca Bailey¹, Jason D. Cooper¹, Gloria Ribas², R. Duncan Campbell^{2,3}, The Wellcome Trust Case Control Consortium†, David G. Clayton¹ & John A. Todd

The major histocompatibility complex (MHC) on chromosome 6 is associated with susceptibility to more common diseases than any other region of the human genome, including almost all disorders classified as autoimmune. In type 1 diabetes the major genetic susceptibility determinants have been mapped to the MHC class II genes *HLA-DQB1* and *HLA-DRB1* (refs 1–3), but these genes cannot completely explain the association between type 1 diabetes and the MHC region^{4–11}. Owing to the region's extreme gene density, the multiplicity of disease-associated alleles, strong associations between alleles, limited genotyping capability, and inadequate statistical approaches and sample sizes, which, and how many, loci within the MHC determine susceptibility remains unclear. Here, in several large type 1 diabetes data sets, we analyse a combined total of 1,729 polymorphisms, and apply statistical methods—recursive partitioning and regression—to pinpoint disease susceptibility to the MHC class I genes *HLA-B* and *HLA-A* (risk ratios >1.5; $P_{\text{combined}} = 2.01 \times 10^{-19}$ and 2.35×10^{-13} , respectively) in addition to the established associations of the MHC class II genes. Other loci with smaller and/or rarer effects might also be involved, but to find these, future searches must take into account both the HLA class II and class I genes and use even larger samples. Taken together with previous studies^{4–8,10–16}, we conclude that MHC-class-I-mediated events, principally involving *HLA-B*39*, contribute to the aetiology of type 1 diabetes.

The MHC spans 4 megabases (Mb) and contains 149 genes, of which eight (the class II loci *HLA-DRB1*, *HLA-DQB1*, *HLA-DQA1*, *HLA-DPB1*, *HLA-DPA1*; the class I loci *HLA-A*, *HLA-B* and *HLA-C*) are the highly polymorphic immune response genes. There are many other candidate genes with common variants—any one of which or a combination thereof—that might also be involved in disease susceptibility. We studied 850 type-1-diabetes-affected sibling-pair (ASP) families from the United Kingdom and the United States and a first set of 2,049 type 1 diabetes patients and 1,912 controls from across Great Britain, in which we genotyped a combined total of 254 polymorphic MHC loci, including *HLA-DQB1*, *HLA-DRB1*, *HLA-A* and *HLA-B* (Table 1 and Supplementary Table 1). A second independent set of 1,050 type 1 diabetes cases and 1,125 controls was used for validation. After these analyses were completed, 1,475 additional single nucleotide polymorphisms (SNPs) in 1,964 of our type 1 diabetes cases and 2,923 controls became available as part of our collaboration with the Wellcome Trust Case Control Consortium (WTCCC)¹⁷ (Table 1).

As expected^{1–3,18}, the strongest type 1 diabetes associations mapped to the MHC class II genes *HLA-DQB1* and *HLA-DRB1* ($P = 10^{-117}$ and $P = 10^{-124}$, respectively, for the genotype model in the families, and similarly, $P < 10^{-300}$ and $P = 10^{-300}$, respectively, in the first case-control set; Fig. 1 and Supplementary Table 2, see also http://dil.tlbase.org/page/poster/mhc_association). The data did not fit the multiplicative model owing to the known epistatic interactions between alleles and dominance effects of *HLA-DRB1* and *HLA-DQB1* genotypes^{2,19}.

There was evidence for a secondary peak of type 1 diabetes association around *HLA-B* ($P = 3.44 \times 10^{-30}$ and 3.59×10^{-42} , in the families and the case-control set, respectively; Fig. 1 and Supplementary Table 2). To test whether these were MHC-class-II-independent effects, or merely reflected linkage disequilibrium with class II, we had to use a method that takes into account the complex multi-allelic effects of the highly disease-associated *HLA-DQB1* and *HLA-DRB1* genes. We compared three strategies for grouping class II genotypes in our families (Supplementary Results). The P -value for the test locus, conditional on the class II genotypes, was, at some loci, dependent on the method adopted for grouping the class II loci (Supplementary Results and Supplementary Table 3). Hence, these methods were unsatisfactory and we adopted a classification tree approach, namely, recursive partitioning^{20–22} (<http://cran.r-project.org>). This is a risk-categorization method of grouping that differs from other risk-based grouping methods because it does not require the risk to be known a priori. The method classifies individuals as affected or unaffected using their class II genotypes by carrying out a series of binary splits on the basis of those class II genotypes, such that homogeneity with respect to disease status (risk) is maximized for each group while retaining good statistical power (see Methods). In contrast to other grouping methods considered, the recursive partitioning model provided consistent results (Supplementary Table 3).

In the families, using the optimized tree model, we found evidence of an additional effect of *HLA-B* ($P = 4.19 \times 10^{-7}$), *HLA-DPB1* ($P = 2.21 \times 10^{-5}$) and of a SNP in the *TAP2* gene (rs241448, $P = 5.29 \times 10^{-5}$; Fig. 1 and Supplementary Table 2). In the case-control set, we again found evidence of an independent effect of *HLA-B* ($P = 1.74 \times 10^{-7}$) over and above the combined effect of *HLA-DQB1* and *HLA-DRB1*, as well as an independent effect of *HLA-A* ($P = 1.67 \times 10^{-10}$). No evidence was obtained for independent effects of the 169 non-synonymous (ns) SNPs, or the seven candidate SNPs ($P > 0.001$, Fig. 1 and Supplementary Table 2). Specifically, no

¹Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge CB2 0XY, UK. ²MRC Rosalind Franklin Centre for Genomics Research, Hinxton, Cambridge CB10 1SB, UK. ³Department of Physiology Anatomy and Genetics, University of Oxford OX1 3QX, UK.

*These authors contributed equally to this work.

†Lists of participants and affiliations appear at the end of the paper.

Table 1 | Data sets used to test for unconditional single locus and MHC-class-II-independent type 1 diabetes associations in the MHC

Data set	Size of data set	Classical MHC genes genotyped in data set	SNPs and microsatellite markers genotyped in data set
Family set	850 affected sibling-pair (ASP) families from the United Kingdom and the United States	<i>HLA-DQB1</i> , <i>HLA-DRB1</i> , <i>HLA-DQA1</i> , <i>HLA-DPB1</i> , <i>HLA-A</i> , <i>HLA-B</i> and <i>HLA-C*</i>	27 microsatellites; 46 SNPs in candidate genes†
First case-control set	2,049 type 1 diabetes patients and 1,912 controls from Great Britain	<i>HLA-DQB1</i> , <i>HLA-DRB1</i> , <i>HLA-A</i> and <i>HLA-B*</i>	7 SNPs in the candidate genes‡; 169 nsSNPs from the GWA scan§
Second case-control set (for validation of the <i>HLA-A</i> association)	Additional 1,050 type 1 diabetes patients and 1,125 controls from Great Britain	<i>HLA-DQB1</i> , <i>HLA-DRB1</i> and <i>HLA-A</i>	Not applicable
WTCCC case-control set	1,964 type 1 diabetes patients and 2,923 controls—overlaps with the main case-control set in 1,281 type 1 diabetes cases and 860 controls	<i>HLA-DQB1</i> , <i>HLA-DRB1</i> , <i>HLA-A</i> and <i>HLA-B</i> : available for 1,281 type 1 diabetes patients and 860 controls	1,475 SNPs from WTCCC GWA scan
WTCCC follow-up case-control set	2,484 type 1 diabetes patients and 2,019 controls	<i>HLA-DQB1</i> , <i>HLA-DRB1</i> , <i>HLA-A</i> and <i>HLA-B</i>	The eight most class-II-independently-associated SNPs from the WTCCC scan

* See Supplementary Table 1.

† Forty-six newly typed SNPs from fifteen candidate immune genes in or near the MHC (*ITPR3*, *HLA-DPB1*, *HLA-DMA*, *HLA-DMB*, *PPP1R2P1*, *TAP1*, *TAP2*, *HLA-DOB*, *BTNL2*, *C6orf25*, *LY6G6C*, *NCR3*, *TNFA*, *LTA* and *NFKBIL1*).

‡ SNPs rs241447 and rs241448 from *TAP2*, rs1800750 from *TNFA*, and rs2296336, rs3131020, rs1233478 and rs389419 in the recently reported type-1-diabetes-associated genes *ITPR3* (ref. 23), *UBD* and *MASIL* (ref. 24).

§ One-hundred-and-sixty-nine nsSNPs were part of a genome-wide association (GWA) scan of over 12,000 nsSNPs³⁰, spanning the entire 10-Mb extended MHC region. These SNPs were analysed in the case-control set and an additional 2,077 type 1 diabetes patients and 2,482 controls from Great Britain that were not typed at the MHC class II loci.

evidence was found for association of the *TAP2* SNP (rs241448, $P = 0.074$), the recently reported *ITPR3* SNP²³ (rs2296336), nor for the *UBD* and *MASIL* gene regions²⁴ (Supplementary Results and Supplementary Table 2).

After conditioning on *HLA-B*, *HLA-DRB1* and *HLA-DQB1*, we had 49% power to find an effect of odds ratio 2.0 in the first case-control set, assuming a minor allele frequency of 0.1 at

$\alpha = 1 \times 10^{-5}$ with *HLA-B*, *HLA-DRB1* and *HLA-DQB1* in the model (Supplementary Methods). Hence, we conditioned on *HLA-B*, *HLA-DRB1* and *HLA-DQB1*, obtaining evidence that *HLA-A* was independently associated with type 1 diabetes ($P = 2.31 \times 10^{-7}$), as was rs4151651 ($P = 8.13 \times 10^{-5}$), a nsSNP in the complement factor B (*CFB*) gene. However, we only had 13% power to test for additional associations to *HLA-B*, *HLA-DRB1* and *HLA-DQB1* in our 850 families, probably accounting for our failure to detect the *HLA-A* association in these families. So we sought to replicate the *HLA-A* result in an independent 1,050 cases and 1,125 controls (Table 1, second case-control set), obtaining convincing confirmatory evidence at $P = 1.77 \times 10^{-5}$ after conditioning on both *HLA-DRB1* and *HLA-DQB1* (Supplementary Table 4).

Having taken into account the combined effect of *HLA-DQB1* and *HLA-DRB1*, as above, we found that the *HLA-B**39 allele (where * represents the allele) was consistently associated with type 1 diabetes susceptibility (relative risk = 3.55 (95% confidence interval 2.21–5.72) in the families; odds ratio = 2.41 (95% confidence interval 1.49–3.89) in the first case-control set; Table 2 and Supplementary Table 5). Moreover, *HLA-B**39 was also associated with a lower age-at-diagnosis of type 1 diabetes in the families ($P = 0.0022$) and in the cases from the first case-control set ($P = 0.0021$; Supplementary Table 5). Once the association of *HLA-B**39 was taken into account, there was no association of other *HLA-B* alleles in the families ($P = 0.047$). Nevertheless, in the first case-control set *HLA-B**18 conferred susceptibility to and *HLA-B**27 protection from type 1 diabetes (Table 2 and Supplementary Table 5). These *HLA-B* allele associations were still present after conditioning on *HLA-A* as well as *HLA-DRB1* and *HLA-DQB1* combined (Table 2).

In the first case-control set, having conditioned on *HLA-DQB1*, *HLA-DRB1* and *HLA-B* using allele *HLA-A**02 as a reference, *HLA-A**01, *HLA-A**11 and *HLA-A**31 were protective and *HLA-A**24 was predisposing for type 1 diabetes; *HLA-A**03 was more predisposing than *HLA-A**11 and *HLA-A**31 (Supplementary Table 4). Once these alleles were accounted for, there was no further detectable *HLA-A* effect in the case-control set ($P = 0.15$). In the second case-control set, having conditioned on *HLA-DRB1* and *HLA-DQB1*, both *HLA-A**01 and *HLA-A**11 were again more protective than *HLA-A**02. *HLA-A**24 was still the most predisposing for type 1 diabetes and may also be associated with an earlier age-at-diagnosis ($P = 0.01$; Supplementary Tables 4 and 5).

Finally, the SNPs from the WTCCC¹⁷ scan were analysed for association with type 1 diabetes. The 20 most associated SNPs all lay within the MHC class II region, with the most associated locus, rs9273363, close to *HLA-DQB1* ($P = 4.29 \times 10^{-298}$ in 1,964 cases and 2,923 controls; Fig. 2 and Supplementary Table 6; see also http://dil.t1dbase.org/page/poster/mhc_association). Once the effects of

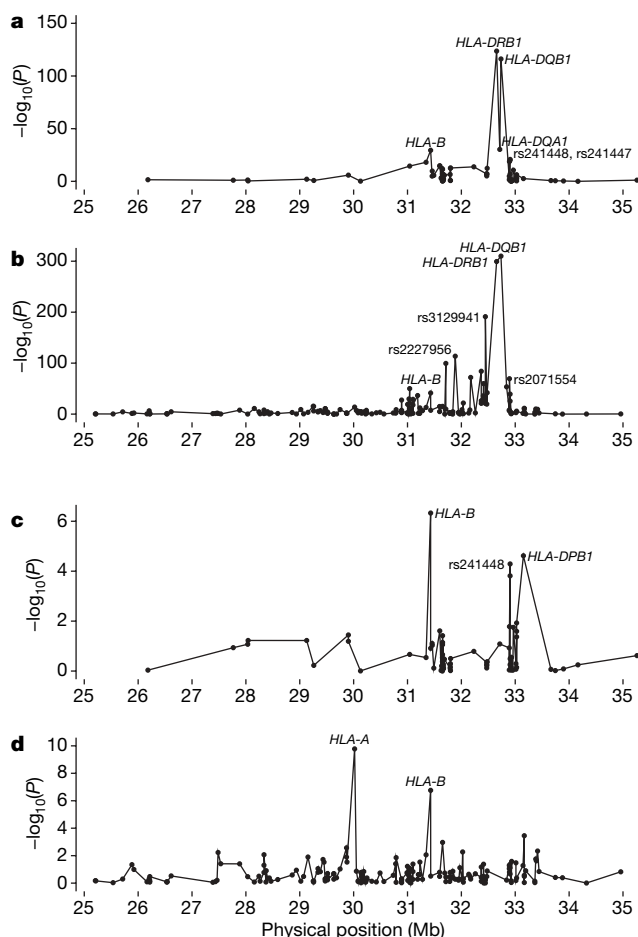


Figure 1 | Association analyses across the MHC. **a**, $-\log_{10}(P)$ versus chromosome position. Unconditional single locus analyses are presented for loci typed in up to 850 families (**a**) and in up to 2,049 cases and 1,125 controls (**b**, first case-control set). **c**, **d**, Analyses conditional on *HLA-DRB1* and *HLA-DQB1* in the families (**c**) and in the first case-control set (**d**). Results are listed in Supplementary Tables 1 and 2.

Table 2 | Type 1 diabetes association of the *HLA-B* alleles (with frequencies >0.015) conditioned on the MHC class II genes and *HLA-A*

HLA-B allele	Allele frequency*		Families (RR (95% CI))		Case-control set (OR (95% CI))	
	Number of cases (%)	Number of controls (%)	Conditioning on <i>HLA-DQB1</i> and <i>HLA-DRB1</i>		Conditioning on <i>HLA-DQB1</i> and <i>HLA-DRB1</i>	
<i>HLA-B*39</i>	143 (4.7)	79 (2.4)	3.55 (2.21–5.72)		2.41 (1.49–3.89)	
<i>HLA-B*18</i>	202 (6.7)	128 (3.8)	1.77 (1.24–2.53)		1.83 (1.19–2.82)	
<i>HLA-B*13</i>	43 (1.4)	56 (1.7)	1.17 (0.65–2.12)		1.94 (0.98–3.85)	
<i>HLA-B*08</i>	788 (26.1)	461 (13.8)	1.26 (0.96–1.65)		0.95 (0.72–1.27)	
<i>HLA-B*55</i>	37 (1.2)	71 (2.1)	0.73 (0.38–1.39)		1.28 (0.64–2.56)	
<i>HLA-B*07</i>	249 (8.2)	465 (13.9)	1.14 (0.84–1.55)		1.25 (0.88–1.78)	
<i>HLA-B*44</i>	344 (11.4)	567 (17.0)	1.00 (reference)		1.00 (reference)	
<i>HLA-B*51</i>	78 (2.6)	109 (3.3)	0.96 (0.61–1.52)		0.93 (0.54–1.62)	
<i>HLA-B*15</i>	388 (12.8)	267 (8.0)	1.24 (0.91–1.69)		0.98 (0.70–1.36)	
<i>HLA-B*35</i>	130 (4.3)	197 (5.9)	1.00 (0.70–1.43)		0.82 (0.55–1.22)	
<i>HLA-B*40</i>	257 (8.5)	232 (6.9)	1.17 (0.86–1.61)		0.87 (0.61–1.22)	
<i>HLA-B*37</i>	21 (0.69)	51 (1.5)	1.03 (0.51–2.08)		0.67 (0.28–1.58)	
<i>HLA-B*14</i>	54 (1.8)	159 (4.8)	1.16 (0.75–1.80)		0.66 (0.39–1.12)	
<i>HLA-B*57</i>	26 (0.86)	143 (4.3)	0.76 (0.37–1.55)		0.50 (0.25–1.00)	
<i>HLA-B*27</i>	113 (3.7)	145 (4.3)	1.02 (0.69–1.52)		0.52 (0.34–0.80)	

Alleles are ordered by risk in the case-control set (once *HLA-DQB1*, *HLA-DRB1* and *HLA-A* have been accounted for). The most common allele, *HLA-B*44*, gives the tightest 95% confidence intervals, so is used as a reference. Results are given for the families (736) and the first case-control set (1,451 type 1 diabetes patients and 1,628 controls) that were successfully typed at all four classical HLA loci. Note that *HLA-B*08* is not a primary effect in type 1 diabetes and is only elevated in frequency in type 1 diabetes cases because of its strong linkage disequilibrium with *HLA-DRB1*03* ($D' = 0.8$). CI, confidence interval; OR, odds ratios; RR, relative risks.

* Allele frequencies in the families are shown in Supplementary Table 1.

HLA-DRB1 and *HLA-DQB1* are accounted for, the polymorphisms in the MHC class I region provide the strongest signals of association in the 1,281 cases and 860 controls genotyped at *HLA-DRB1*, *HLA-DQB1* and the WTCCC SNPs (Fig. 2 and Supplementary Table 6). The most associated locus became rs3130531, located ~40 kilobases telomeric of *HLA-C* ($P = 6.74 \times 10^{-7}$ compared with $P = 0.0056$ before conditioning). Once *HLA-DRB1*, *HLA-DQB1* and *HLA-B* were conditioned on, however, none of the WTCCC SNPs were convincingly associated with type 1 diabetes (Fig. 2 and http://dil.tlbase.org/page/poster/mhc_association).

Nevertheless, to increase our statistical power, we genotyped the eight WTCCC SNPs most associated with type 1 diabetes, after *HLA-DRB1* and *HLA-DQB1* conditioning, in a larger set of 2,484 cases and 2,019 controls with complete *HLA-DRB1* and *HLA-DQB1* genotyping. The most associated locus was rs9268831 ($P = 6.95 \times 10^{-8}$; Supplementary Table 6); note that this was less significant than either *HLA-B* or *HLA-A*, which remained the most associated loci in the same data set after class II genes were accounted for ($P = 3.80 \times 10^{-17}$ and 4.59×10^{-15} , respectively; Supplementary Table 6). This SNP, located ~15 kb centromeric of *HLA-DRA*,

was still associated once *HLA-B* was included in the model ($P = 5.44 \times 10^{-6}$). In contrast, the SNP rs3130531 at 31.3 Mb was not associated in this data set after conditioning on *HLA-DRB1*, *HLA-DQB1* and *HLA-B* ($P = 0.16$; Supplementary Table 6).

Our results indicate that, once the effect of the MHC class II genes has been accounted for, most of the detectable residual association is attributable to *HLA-B* and *HLA-A* (combining all data sets, $P_{\text{combined}} = 2.01 \times 10^{-19}$ and 2.35×10^{-13} , respectively). We conclude that the existence of other major type 1 diabetes genes in the extended MHC is unlikely. Smaller independent effects, however, might still exist, necessitating future studies including analysis of rs9268831 (*HLA-DRA*), rs4151651 (*CFB*), *HLA-C*, *HLA-DQA1* and the *HLA-DP* loci. The *HLA-B* and *HLA-A* alleles have previously been associated with type 1 diabetes^{4,5,8,10,11}, but unlike these previous studies, our results localize the effects to these specific loci and alleles, thereby implicating them directly in disease aetiology.

In the nonobese diabetic mouse model of type 1 diabetes, MHC class I molecules and class-I-restricted CD8⁺ T cells are central to the development of autoimmune diabetes^{12–14}. This correlates with the observations that in type 1 diabetes patients, cells infiltrating

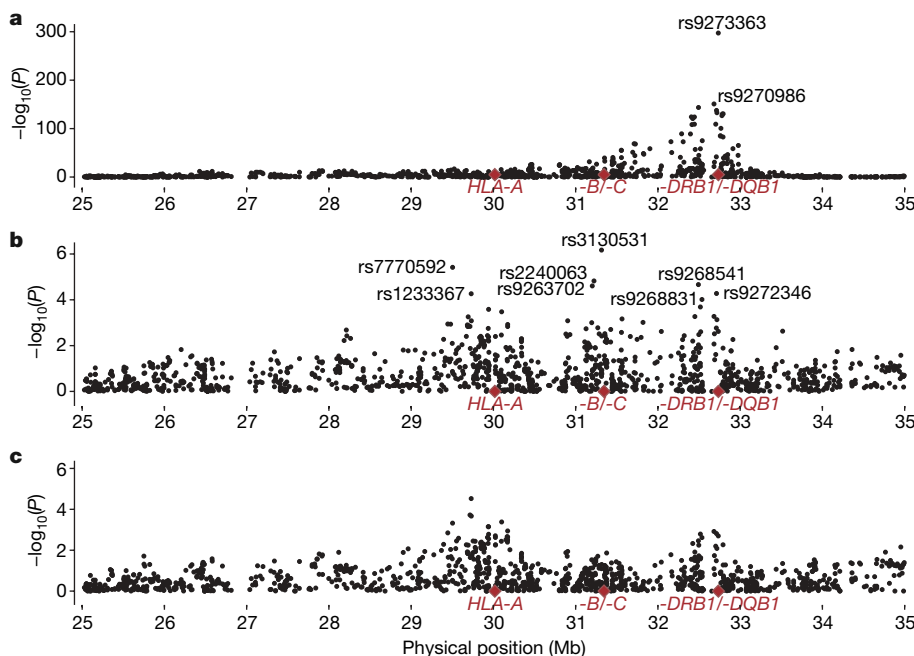


Figure 2 | Association analyses of 1,475 SNPs across the MHC. a, Unconditional single locus analysis in up to 1,964 cases and 2,923 controls—the limits of the association are at 25.9 Mb (rs1324088 $P = 4.65 \times 10^{-6}$) and 34.0 Mb (rs6941621 $P = 9.95 \times 10^{-6}$). **b**, **c**, Results are presented in up to 1,281 cases and 860 controls for analyses conditioned on *HLA-DRB1* and *HLA-DQB1* combined (annotated SNPs were followed up in a larger case-control set; Supplementary Table 6) (**b**), and for analyses conditioned on *HLA-DRB1* and *HLA-DQB1* combined and the alleles of *HLA-B* with frequency >0.01 (**c**).

pancreatic islets are predominantly CD8⁺ and islet cells hyperexpress MHC class I molecules^{15,16}. Taken together with our results, we conclude that class-I-mediated anti-islet β -cell responses are critical in type 1 diabetes and may accelerate disease onset. This might involve both the innate and adaptive immune system^{25,26}. The HLA-A*02 allotype has been functionally and directly linked to T-cell autoreactivity to insulin²⁷, and our results now justify investigation of the naturally processed peptides that bind the HLA-B*39 allotype as a first step towards future evaluation of inducing tolerance to such peptides in attempts to prevent type 1 diabetes.

METHODS SUMMARY

A detailed description of the methods is given in Methods and Supplementary Information. All subjects were of white ethnicity: 850 families came from established collections; 4,126 cases were Juvenile Diabetes Research Foundation/Wellcome Trust type 1 diabetes cases and 4,394 were British 1958 Birth Cohort controls. The classical loci were typed using Dynal RELI SSO assays. All loci conformed to Hardy–Weinberg equilibrium in unaffected subjects. Family data were analysed using cases with matched pseudo-controls in regression models. Stepwise logistic regression²⁸ was used to test for effects independent of HLA-DRB1 and HLA-DQB1. The class II genotypes (rather than the alleles of HLA-DRB1 and HLA-DQB1 which do not behave multiplicatively in conferring type 1 diabetes risk) were modelled using a recursive partitioning approach^{20–22} (<http://cran.r-project.org>). These groups created by recursive partitioning, define strata within which additional loci can be tested. Pruning of the tree, that is, assessing how much of the tree to retain, was done by cross-validation²⁹. Although when to stop pruning the tree was unclear, the number of leaves (HLA-DQB1/HLA-DRB1 groups) must be sufficient to prevent residual confounding, but not so great as to become inestimable. The recursive partitioning model was assessed using a third locus. The non-class-II loci were modelled as multiplicative effects of alleles. All analyses of the case–control set were stratified by broad geographical region^{17,30}. Power to detect effects independent of HLA-DRB1 and HLA-DQB1 was assessed using data sets simulated from the case–pseudo-control data (and separately from the case–control set). Crucially the groupings from the recursive partitioning model were retained. A total of 100,000 replicates were performed. The Wald test was used to assess significance and subsequently the power of the study. We had 49% power to detect odds ratio = 2.0 for an allele frequency of 0.1 at $\alpha = 1 \times 10^{-5}$ in the case–control set with HLA-DRB1, HLA-DQB1 and HLA-B in the model. We set a threshold of $P < 0.0001$ for rejection of H₀.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 July; accepted 25 October 2007.

Published online 14 November 2007.

- Todd, J. A., Bell, J. I. & McDevitt, H. O. HLA-DQ β gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* **329**, 599–604 (1987).
- Cucca, F. *et al.* A correlation between the relative predisposition of MHC class II alleles to type 1 diabetes and the structure of their proteins. *Hum. Mol. Genet.* **10**, 2025–2037 (2001).
- Todd, J. A. *et al.* Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nature Genet.* **39**, 857–864 (2007).
- Fennessy, M. *et al.* A gene in the HLA class I region contributes to susceptibility to IDDM in the Finnish population. Childhood Diabetes in Finland (DiMe) Study Group. *Diabetologia* **37**, 937–944 (1994).
- Nejentsev, S. *et al.* Non-class II HLA gene associated with type 1 diabetes maps to the 240-kb region near HLA-B. *Diabetes* **49**, 2217–2221 (2000).
- Lie, B. A. *et al.* The predisposition to type 1 diabetes linked to the human leukocyte antigen complex includes at least one non-class II gene. *Am. J. Hum. Genet.* **64**, 793–800 (1999).
- Valdes, A. M. *et al.* Extended DR3–D6S273–HLA-B haplotypes are associated with increased susceptibility to type 1 diabetes in US Caucasians. *Tissue Antigens* **65**, 115–119 (2005).
- Valdes, A. M., Erlich, H. A. & Noble, J. A. Human leukocyte antigen class I B and C loci contribute to Type 1 Diabetes (T1D) susceptibility and age at T1D onset. *Hum. Immunol.* **66**, 301–313 (2005).
- Aly, T. A. *et al.* Extreme genetic risk for type 1A diabetes. *Proc. Natl Acad. Sci. USA* **103**, 14074–14079 (2006).
- Noble, J. A. *et al.* The HLA class I A locus affects susceptibility to type 1 diabetes. *Hum. Immunol.* **63**, 657–664 (2002).
- Honeyman, M. C., Harrison, L. C., Drummond, B., Colman, P. G. & Tait, B. D. Analysis of families at risk for insulin-dependent diabetes mellitus reveals that HLA antigens influence progression to clinical disease. *Mol. Med.* **1**, 576–582 (1995).

- Wicker, L. S. *et al.* β 2-microglobulin-deficient NOD mice do not develop insulinitis or diabetes. *Diabetes* **43**, 500–504 (1994).
- Utsugi, T. *et al.* Major histocompatibility complex class I-restricted infiltration and destruction of pancreatic islets by NOD mouse-derived beta-cell cytotoxic CD8⁺ T-cell clones *in vivo*. *Diabetes* **45**, 1121–1131 (1996).
- Marron, M. P., Graser, R. T., Chapman, H. D. & Serreze, D. V. Functional evidence for the mediation of diabetogenic T cell responses by HLA-A2.1 MHC class I molecules through transgenic expression in NOD mice. *Proc. Natl Acad. Sci. USA* **99**, 13753–13758 (2002).
- Itoh, N. *et al.* Mononuclear cell infiltration and its relation to the expression of major histocompatibility complex antigens and adhesion molecules in pancreas biopsy specimens from newly diagnosed insulin-dependent diabetes mellitus patients. *J. Clin. Invest.* **92**, 2313–2322 (1993).
- Foulis, A. K., Liddle, C. N., Farquharson, M. A., Richmond, J. A. & Weir, R. S. The histopathology of the pancreas in type 1 (insulin-dependent) diabetes mellitus: a 25-year review of deaths in patients under 20 years of age in the United Kingdom. *Diabetologia* **29**, 267–274 (1986).
- Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
- Herr, M. *et al.* Evaluation of fine mapping strategies for a multifactorial disease locus: systematic linkage and association analysis of *IDDM1* in the HLA region on chromosome 6p21. *Hum. Mol. Genet.* **9**, 1291–1301 (2000).
- Noble, J. A. *et al.* The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. *Am. J. Hum. Genet.* **59**, 1134–1148 (1996).
- Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. *Classification and Regression Trees* (Wadsworth, Belmont, 1984).
- R Development Core Team. *A Language and Environment for Statistical Computing*. (R Foundation for statistical computing, Vienna, 2006).
- Therneau, T. M. & Atkinson, E. J. *An Introduction to Recursive Partitioning Using the rpart Routine*. Technical report no. 61 (Mayo Clinic, section of statistics, Minnesota, 1997).
- Roach, J. C. *et al.* Genetic mapping at 3-kilobase resolution reveals inositol 1,4,5-triphosphate receptor 3 as a risk factor for type 1 diabetes in Sweden. *Am. J. Hum. Genet.* **79**, 614–627 (2006).
- Aly, T. A. *et al.* High density SNP analysis of the MHC region reveals multiple loci for type 1A diabetes. *Clin. Immunol.* **123**, S133 (2007).
- de Jersey, J. *et al.* Beta cells cannot directly prime diabetogenic CD8 T cells in nonobese diabetic mice. *Proc. Natl Acad. Sci. USA* **104**, 1295–1300 (2007).
- Poirot, L., Benoist, C. & Mathis, D. Natural killer cells distinguish innocuous and destructive forms of pancreatic islet autoimmunity. *Proc. Natl Acad. Sci. USA* **101**, 8102–8107 (2004).
- Pinkse, G. G. *et al.* Autoreactive CD8 T cells associated with beta cell destruction in type 1 diabetes. *Proc. Natl Acad. Sci. USA* **102**, 18425–18430 (2005).
- Cordell, H. J. & Clayton, D. G. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am. J. Hum. Genet.* **70**, 124–141 (2002).
- Stone, M. Cross-validation choice and assessment of statistical predictions. *J. R. Stat. Soc. B* **36**, 111–147 (1974).
- Clayton, D. G. *et al.* Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nature Genet.* **37**, 1243–1246 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was funded by the Wellcome Trust and the Juvenile Diabetes Research Foundation International. We thank all of the patients, control subjects and family members for their participation. The Human Biological Data Interchange and Diabetes UK Warren repositories and UK GRID project are acknowledged for the collection of the type 1 diabetes patients and families. We acknowledge use of DNA from the British 1958 Birth Cohort collection (D. Strachan, S. Ring, W. McArdle, P. Burton, R. Jones and M. Pembrey), funded by the Medical Research Council and Wellcome Trust. S.N. is a Diabetes Research and Wellness Foundation Non-Clinical Fellow. R.D.C. and G.R. were funded by the Medical Research Council.

Author Contributions J.M.M.H. performed all statistical data analyses, interpreted results, contributed to the direction of the study and wrote the manuscript. S.N. participated in the conception, design and coordination of the study, genotyping, data analysis and writing of the manuscript. J.A.T. participated in the conception, design and coordination of the study, as well as data analysis and writing of the manuscript. N.M.W. curated the data and helped coordinate the HLA typing. H.E.S. was responsible for DNA. J.S., S.F.F., P.R., M.H., E.K., J.M., J.H., L.M.M., D.S. and R.B. contributed to genotyping of SNPs, microsatellites and HLA loci. J.D.C. provided nsSNP GWA data. G.R. provided SNPs and genotyping in candidate genes from the MHC class III genes. R.D.C. provided SNPs in candidate genes from MHC class III genes. The Wellcome Trust Case Control Consortium provided GWA SNP data. D.G.C. gave guidance on statistical analyses.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.M.M.H. (Joanna.Howson@cimr.cam.ac.uk) or J.A.T. (John.Todd@cimr.cam.ac.uk).

The Wellcome Trust Case Control Consortium

Management committee Paul R. Burton¹, David G. Clayton², Lon R. Cardon³, Nick Craddock⁴, Panos Deloukas⁵, Audrey Duncanson⁶, Dominic P. Kwiatkowski^{3,5}, Mark L. McCarthy^{3,7}, Willem H. Ouwehand^{8,9}, Nilesh J. Samani¹⁰, John A. Todd² & Peter Donnelly (Chair)¹¹

Analysis committee Jeffrey C. Barrett³, Paul R. Burton¹, Dan Davison¹¹, Peter Donnelly¹¹, Doug Easton¹², David Evans³, Hin-Tak Leung², Jonathan L. Marchini¹¹, Andrew P. Morris³, Chris C. A. Spencer¹¹, Martin D. Tobin¹, Lon R. Cardon (Co-chair)³ & David G. Clayton (Co-chair)²

UK blood services and University of Cambridge controls Antony P. Attwood^{5,8}, James P. Boorman^{8,9}, Barbara Cant⁸, Ursula Everson¹³, Judith M. Hussey¹⁴, Jennifer D. Jolley⁵, Alexandra S. Knight⁵, Kerstin Koch⁸, Elizabeth Meech¹⁵, Sarah Nutland², Christopher V. Prowse¹⁶, Helen E. Stevens², Niall C. Taylor⁸, Graham R. Walters¹⁷, Neil M. Walker², Nicholas A. Watkins^{8,9}, Thilo Winzer⁵, John A. Todd² & Willem H. Ouwehand^{8,9}

1958 birth cohort controls Richard W. Jones¹⁸, Wendy L. McArdle¹⁸, Susan M. Ring¹⁸, David P. Strachan¹⁹ & Marcus Pembrey^{18,20}

Bipolar disorder Gerome Breen²¹, David St Clair²¹ (Aberdeen); Sian Caesar²², Katherine Gordon-Smith^{22,23}, Lisa Jones²² (Birmingham); Christine Fraser²³, Elaine K. Green²³, Detelina Grozeva²³, Marian L. Hamshere²³, Peter A. Holmans²³, Ian R. Jones²³, George Kirov²³, Valentina Moskvina²³, Ivan Nikolov²³, Michael C. O'Donovan²³, Michael J. Owen²³, Nick Craddock²³ (Cardiff); David A. Collier²⁴, Amanda Elkin²⁴, Anne Farmer²⁴, Richard Williamson²⁴, Peter McGuffin²⁴ (London); Allan H. Young²⁵ & I. Nicol Ferrier²⁵ (Newcastle)

Coronary artery disease Stephen G. Ball²⁶, Anthony J. Balmforth²⁶, Jennifer H. Barrett²⁶, D. Timothy Bishop²⁶, Mark M. Iles²⁶, Azhar Maqbool²⁶, Nadira Yuldasheva²⁶, Alistair S. Hall²⁶ (Leeds); Peter S. Braund¹⁰, Paul R. Burton¹, Richard J. Dixon¹⁰, Massimo Mangino¹⁰, Suzanne Stevens¹⁰, Martin D. Tobin¹, John R. Thompson¹ & Nilesh J. Samani¹⁰ (Leicester)

Crohn's disease Francesca Bredin²⁷, Mark Tremelling²⁷, Miles Parkes²⁷ (Cambridge); Hazel Drummond²⁸, Charles W. Lees²⁸, Elaine R. Nimmo²⁸, Jack Satsangi²⁸ (Edinburgh); Sheila A. Fisher²⁹, Alastair Forbes³⁰, Cathryn M. Lewis²⁹, Clive M. Onnie²⁹, Natalie J. Prescott²⁹, Jeremy Sanderson³¹, Christopher G. Mathew²⁹ (London); Jamie Barbour³², M. Khalid Mohiuddin³², Catherine E. Todhunter³², John C. Mansfield³² (Newcastle); Tariq Ahmad³³, Fraser R. Cummings³³ & Derek P. Jewell³³ (Oxford)

Hypertension John Webster³⁴ (Aberdeen); Morris J. Brown³⁵, David G. Clayton² (Cambridge); G. Mark Lathrop³⁶ (Evry, France); John Connell³⁷, Anna Dominiczak³⁷ (Glasgow); Nilesh J. Samani¹⁰ (Leicester); Carolina A. Braga Marciano³⁸, Beverley Burke³⁸, Richard Dobson³⁸, Johannie Gungadoo³⁸, Kate L. Lee³⁸, Patricia B. Munroe³⁸, Stephen J. Newhouse³⁸, Abiodun Onipinla³⁸, Chris Wallace³⁸, Mingzhan Xue³⁸, Mark Caulfield³⁸ (London); Martin Farrall³⁹ (Oxford)

Rheumatoid arthritis Anne Barton⁴⁰, The Biologics in RA Genetics and Genomics Study Syndicate (BRAGGS) Steering Committee*, Ian N. Bruce⁴⁰, Hannah Donovan⁴⁰, Steve Eyre⁴⁰, Paul D. Gilbert⁴⁰, Samantha L. Hider⁴⁰, Anne M. Hinks⁴⁰, Sally L. John⁴⁰, Catherine Potter⁴⁰, Alan J. Silman⁴⁰, Deborah P. M. Symmons⁴⁰, Wendy Thomson⁴⁰ & Jane Worthington⁴⁰

Type 1 diabetes David G. Clayton², David B. Dunger^{2,41}, Sarah Nutland², Helen E. Stevens², Neil M. Walker², Barry Widmer^{2,41} & John A. Todd²

Type 2 diabetes Timothy M. Frayling^{42,43}, Rachel M. Freathy^{42,43}, Hana Lango^{42,43}, John R. B. Perry^{42,43}, Beverley M. Shields⁴³, Michael N. Weedon^{42,43}, Andrew T. Hattersley^{42,43} (Exeter); Graham A. Hitman⁴⁴ (London); Mark Walker⁴⁵ (Newcastle); Kate S. Elliott^{3,7}, Christopher J. Groves⁷, Cecilia M. Lindgren^{3,7}, Nigel W. Rayner^{3,7}, Nicholas J. Timpson^{3,46}, Eleftheria Zeggini^{3,7} & Mark L. McCarthy^{3,7} (Oxford)

Tuberculosis Melanie Newport⁴⁷, Giorgio Sirugo⁴⁷ (Gambia); Emily Lyons³, Fredrik Vannberg³ & Adrian V. S. Hill³ (Oxford)

Ankylosing spondylitis Linda A. Bradbury⁴⁸, Claire Farrar⁴⁹, Jennifer J. Pointon⁴⁸, Paul Wordsworth⁴⁹ & Matthew A. Brown^{48,49}

Autoimmune thyroid disease Jayne A. Franklyn⁵⁰, Joanne M. Heward⁵⁰, Matthew J. Simmonds⁵⁰ & Stephen C. L. Gough⁵⁰

Breast cancer Sheila Seal⁵¹, Breast Cancer Susceptibility Collaboration (UK)*, Michael R. Stratton^{51,52} & Nazneen Rahman⁵¹

Multiple sclerosis Maria Ban⁵³, An Goris⁵³, Stephen J. Sawcer⁵³ & Alastair Compston⁵³

Gambian controls David Conway⁴⁷, Muminatou Jallow⁴⁷, Melanie Newport⁴⁷, Giorgio Sirugo⁴⁷ (Gambia); Kirk A. Rockett³ & Dominic P. Kwiatkowski^{3,5} (Oxford)

DNA, genotyping, data QC and informatics Claire Bryan⁵, Suzannah J. Bumpstead⁵, Amy Chaney⁵, Kate Downes^{2,5}, Jilur Ghoris⁵, Rhian Gwilliam⁵, Sarah E. Hunt⁵, Michael Inouye⁵, Andrew Keniry⁵, Emma King⁵, Ralph McGinnis⁵, Simon Potter⁵, Rathi Ravindrarajah⁵, Pamela Whittaker⁵, David Withers⁵, Panos Deloukas⁵ (Wellcome Trust Sanger Institute, Hinxton); Hin-Tak Leung², Sarah Nutland², Helen E. Stevens², Neil M. Walker² & John A. Todd² (Cambridge)

Statistics Doug Easton¹², David G. Clayton² (Cambridge); Paul R. Burton¹, Martin D. Tobin¹ (Leicester); Jeffrey C. Barrett³, David Evans³, Andrew P. Morris³, Lon R. Cardon³, Niall J. Cardin¹¹, Dan Davison¹¹, Teresa Ferreira¹¹, Joanne Pereira-Gale¹¹, Ingeleif B. Hallgrimsdottir¹¹, Bryan N. Howie¹¹, Jonathan L. Marchini¹¹, Chris C. A. Spencer¹¹, Zhan Su¹¹, Yik Ying Teo^{3,11}, Damjan Vukcevic¹¹ & Peter Donnelly¹¹ (Oxford)

Primary investigators David Bentley^{5†}, Matthew A. Brown^{48,49}, Lon R. Cardon³, Mark Caulfield³⁸, David G. Clayton², Alistair Compston⁵³, Nick Craddock²³, Panos Deloukas⁵, Peter Donnelly¹¹, Martin Farrall³⁹, Stephen C. L. Gough⁵⁰, Alistair S. Hall²⁶, Andrew T. Hattersley^{42,43}, Adrian V. S. Hill³, Dominic P. Kwiatkowski^{3,5}, Christopher G. Mathew²⁹, Mark L. McCarthy^{3,7}, Willem H. Ouwehand^{8,9}, Miles Parkes²⁷, Marcus Pembrey^{18,20}, Nazneen Rahman⁵¹, Nilesh J. Samani¹⁰, Michael R. Stratton^{51,52}, John A. Todd² & Jane Worthington⁴⁰

*See Supplementary Information for details.

¹Genetic Epidemiology Group, Department of Health Sciences, University of Leicester, Adrian Building, University Road, Leicester LE1 7RH, UK. ²Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Wellcome Trust/MRC Building, Cambridge CB2 0XY, UK. ³Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK.

⁴Department of Psychological Medicine, Henry Wellcome Building, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK. ⁵The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. ⁶The Wellcome Trust, Gibbs Building, 215 Euston Road, London NW1 2BE, UK. ⁷Oxford Centre for Diabetes, Endocrinology and Medicine, University of Oxford, Churchill Hospital, Oxford OX3 7LJ, UK. ⁸Department of Haematology, University of Cambridge, Long Road, Cambridge CB2 2PT, UK. ⁹National Health Service Blood and Transplant, Cambridge Centre, Long Road, Cambridge CB2 2PT, UK. ¹⁰Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Groby Road, Leicester LE3 9QP, UK.

¹¹Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK. ¹²Cancer Research UK Genetic Epidemiology Unit, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 8RN, UK. ¹³National Health Service Blood and Transplant, Sheffield Centre, Longley Lane, Sheffield S5 7JN, UK. ¹⁴National Health Service Blood and Transplant, Brentwood Centre, Crescent Drive, Brentwood CM15 8DP, UK. ¹⁵The Welsh Blood Service, Ely Valley Road, Talbot Green, Pontyclun CF72 9WB, UK. ¹⁶The Scottish National Blood Transfusion Service, Ellen's Glen Road, Edinburgh EH17 7QT, UK. ¹⁷National Health Service Blood and Transplant, Southampton Centre, Coxford Road, Southampton SO16 5AF, UK. ¹⁸Avon Longitudinal Study of Parents and Children, University of Bristol, 24 Tyndall Avenue, Bristol BS8 1TQ, UK. ¹⁹Division of Community Health Services, St George's University of London, Cranmer Terrace, London SW17 0RE, UK. ²⁰Institute of Child Health, University College London, 30 Guilford Street, London WC1N 1EH, UK. ²¹University of Aberdeen, Institute of Medical Sciences, Foresterhill, Aberdeen AB25 2ZD, UK. ²²Department of Psychiatry, Division of Neuroscience, Birmingham University, Birmingham B15 2QZ, UK. ²³Department of Psychological Medicine, Henry Wellcome Building, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK. ²⁴SGDP, The Institute of Psychiatry, King's College London, De Crespigny Park Denmark Hill, London SE5 8AF, UK. ²⁵School of Neurology, Neurobiology and Psychiatry, Royal Victoria Infirmary, Queen Victoria Road, Newcastle upon Tyne NE1 4LP, UK. ²⁶LIGHT and LIMM Research Institutes, Faculty of Medicine and Health, University of Leeds, Leeds LS1 3EX, UK. ²⁷IBD Research Group, Addenbrooke's Hospital, University of Cambridge, Cambridge CB2 2QQ, UK. ²⁸Gastrointestinal Unit, School of Molecular and Clinical Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK. ²⁹Department of Medical & Molecular Genetics, King's College London School of Medicine, 8th Floor Guy's Tower, Guy's Hospital, London SE1 9RT, UK. ³⁰Institute for Digestive Diseases, University College London Hospitals Trust, London NW1 2BU, UK. ³¹Department of Gastroenterology, Guy's and St Thomas' NHS Foundation Trust, London SE1 7EH, UK. ³²Department of Gastroenterology & Hepatology, University of Newcastle upon Tyne, Royal Victoria Infirmary, Newcastle upon Tyne NE1 4LP, UK. ³³Gastroenterology Unit, Radcliffe Infirmary, University of Oxford, Oxford OX2 6HE, UK. ³⁴Medicine and Therapeutics, Aberdeen Royal Infirmary, Foresterhill, Aberdeen, Grampian AB9 2ZB, UK. ³⁵Clinical Pharmacology Unit and the Diabetes and Inflammation Laboratory, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, UK. ³⁶Centre National de Genotypage, 2, Rue Gaston Cremieux, Evry, Paris 91057, France. ³⁷BHF Glasgow Cardiovascular Research Centre, University of Glasgow, 126 University Place, Glasgow G12 8TA, UK. ³⁸Clinical Pharmacology and Barts and The London Genome Centre, William Harvey Research Institute, Barts and The London, Queen Mary's School of Medicine, Charterhouse

Square, London EC1M 6BQ, UK. ³⁹Cardiovascular Medicine, University of Oxford, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK. ⁴⁰arc Epidemiology Research Unit, University of Manchester, Stopford Building, Oxford Road, Manchester M13 9PT, UK. ⁴¹Department of Paediatrics, University of Cambridge, Addenbrooke's Hospital, Cambridge CB2 2QQ, UK. ⁴²Genetics of Complex Traits, Institute of Biomedical and Clinical Science, Peninsula Medical School, Magdalen Road, Exeter EX1 2LU, UK. ⁴³Diabetes Genetics, Institute of Biomedical and Clinical Science, Peninsula Medical School, Barrack Road, Exeter EX2 5DU, UK. ⁴⁴Centre for Diabetes and Metabolic Medicine, Barts and The London, Royal London Hospital, Whitechapel, London E1 1BB, UK. ⁴⁵Diabetes Research Group, School of Clinical Medical Sciences, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, UK. ⁴⁶The MRC Centre for Causal Analyses in Translational Epidemiology, Bristol University,

Canynges Hall, Whiteladies Road, Bristol BS2 8PR, UK. ⁴⁷MRC Laboratories, Fajara, The Gambia. ⁴⁸Diamantina Institute for Cancer, Immunology and Metabolic Medicine, Princess Alexandra Hospital, University of Queensland, Woolloongabba, Queensland 4102, Australia. ⁴⁹Botnar Research Centre, University of Oxford, Headington, Oxford OX3 7BN, UK. ⁵⁰Department of Medicine, Division of Medical Sciences, Institute of Biomedical Research, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. ⁵¹Section of Cancer Genetics, Institute of Cancer Research, 15 Cotswold Road, Sutton SM2 5NG, UK. ⁵²Cancer Genome Project, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. ⁵³Department of Clinical Neurosciences, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, UK. †Present address: Illumina Cambridge, Chesterford Research Park, Little Chesterford, Nr Saffron Walden, Essex CB10 1XL, UK.

METHODS

Subjects. The family set comprised 850 type 1 diabetes families of white ethnicity with both parents and at least two affected children in each family, including 472 Warren type 1 diabetes families from the United Kingdom (11 with only one affected offspring) and 378 Human Biological Data Interchange (HBDI) type 1 diabetes families from the United States (five with only one affected offspring). The case-control sample comprised 4,126 type 1 diabetes patients collected as part of the JDREF/WT DIL British type 1 diabetes case collection (<http://www-gene.cimr.cam.ac.uk/ucdr/grid.shtml>) and 4,394 controls selected from the British 1958 Birth Cohort of people born in England, Scotland and Wales during 1 week in 1958 (<http://www.b58cgenegsugl.ac.uk>). Of these individuals, 2,049 cases and 1,912 controls were typed at the classical MHC loci. A further 1,445 controls were from the WTCCC's UK Blood Service samples¹⁷. The relevant research ethics committees approved the study, and written informed consent was obtained from the participants, or their parents/guardian for those too young to consent.

Grouping of *HLA-DRB1* and *HLA-DQB1* alleles and genotypes: recursive partitioning. In the families, cases and matched pseudo-controls were generated²⁸ and the matching discarded so as to be able to run recursive partitioning in the recursive partitioning library in R (<http://cran.r-project.org>; refs 20–22). Cases and pseudo-controls were matched when doing later analyses. The alleles of *HLA-DRB1* and *HLA-DQB1* do not behave multiplicatively in conferring type 1 diabetes risk, so, to allow for dominance effects, the genotypes of the individual loci were modelled, as this does not assume a specific mode of inheritance. All possible binary splits of the data corresponding to presence or absence of various different genotypes at *HLA-DRB1* and *HLA-DQB1* were considered. The split that best categorized the data as cases and pseudo-controls, which corresponds to the split that maximizes the reduction in impurity (or maximizes the homogeneity of the cases or pseudo-controls within groups), was chosen. We chose the information index, which has the form $f(p) = -2\log(p)$ (where p = the proportion of observations in a node that for future samples belong to a different class) as the impurity measure because it is likelihood based. This process was repeated until no further improvement could be made or the minimum group size was met. The terminal 'leaves' of the tree represent optimized groups of the *HLA-DQB1/HLA-DRB1* genotypes and so define strata within which additional MHC loci can be tested. However, the trees generated are often complex and need to be pruned. Pruning of the tree, that is, assessing how much of the tree to retain, was done by cross-validation²⁹. Nevertheless, when to stop pruning the tree was unclear, the number of leaves (*HLA-DQB1/DRB1* groups) must be sufficient to prevent residual confounding, but not so great as to become inestimable.

Pruning and evaluation of the *HLA-DQB1/HLA-DRB1* trees. Initially, we used the *TAP2* SNP rs241448 to assess the appropriateness of the *HLA-DRB1* and *HLA-DQB1* tree models and subsequently used an additional 14 loci (Supplementary Table 3). As the number of terminal leaves (*HLA-DRB1/HLA-DQB1* groups) in the model increases, we expect the effect size of the additional locus to decrease as confounding is reduced, while the 95% confidence interval will become larger (Supplementary Fig. 1). This information can be used to optimize the choice of tree by considering the best compromise between number of terminal leaves (groups of *HLA-DRB1* and *HLA-DQB1* genotypes), effect size and 95% confidence interval. By adding the *TAP2* SNP rs241448, we tested each of the possible six pruned *HLA-DRB1/HLA-DQB1* trees (that corresponded to different complexity parameters) in the family data set. The greatest disparity in effect size was seen between the two models with the minimum number of terminal leaves (four and seven groups) with relative risk = 0.7 (95% confidence interval 0.5–0.9) and the remainder with between 12 and 22 groups, relative risk = 0.6 (95% confidence interval 0.5–0.8). The effect size and 95% confidence interval at this locus was stable with respect to the number of terminal leaves (that is, *HLA-DRB1/HLA-DQB1* groups). Therefore, using the *TAP2* SNP only, the 12 group model appeared to be the best compromise between effect size, 95% confidence interval and number of groups (that is, complexity of the model). Nevertheless, to verify that the number of groups used for the MHC class II model would not affect the interpretation of results, we tested the remaining 14 loci used to evaluate other grouping methods (Supplementary Results) for association, conditioning on class II effects using each of the four tree models with 12 or greater terminal leaves (Supplementary Table 3). The 12 group model exhibited one inconsistent result at *MICA* compared to all the other recursive partitioning models ($P = 0.0005$ versus $P > 0.05$), otherwise all loci tested were stable with respect to the number of groups in the model (Supplementary Table 3). Hence, the model with 16 terminal leaves was chosen to model the confounding

effects of *HLA-DRB1* and *HLA-DQB1*: this model had the minimum number of groups that gave results consistent with both the 18 group model and the 22 group model at all loci.

We did not use this 16 group model for the case-control collection because the way in which the two sample sets were ascertained could affect their MHC associations. The ASP families are likely to be enriched for HLA susceptibility haplotypes compared to isolated cases. They were also collected over 10 years earlier, during which time the incidence of type 1 diabetes has increased, and have a higher average age-at-diagnosis (12 years) compared to the British cases (7 years). We did, however, use the same approach to choosing the optimal tree model for the cases and controls as for the families. The MHC class II genotypes were put into the recursive partitioning library. The maximum number of groups obtained was 14 (which corresponded to a complexity parameter of 0). Of the five pruned *HLA-DRB1/HLA-DQB1* trees possible (corresponding to different complexity parameters) the model with 12 terminal leaves was the best compromise between effect size and 95% confidence interval, gave consistent results across loci and was very similar to the tree with the maximum 14 terminal leaves.

The effect sizes with corresponding 95% confidence intervals are given in Supplementary Table 7 for the case-control model and the model used for the families, using an approximately neutral group as reference. Note that although the tree used for the families has 16 groups, one of these groups only contains pseudo-controls and so is not used for the analysis. Similarly one of the 12 groups used for the case-control set consisted of just 13 cases, which are dropped from the association analysis.

We then assessed the effectiveness of the *HLA-DRB1/HLA-DQB1* tree model. Hence, we generated 1,000 bootstrap sample data sets, with replacement, within geographical and case-control strata. Four loci were used as the non-class-II test locus and analysed in each data set: the *TAP2* SNP rs241448; the *UBD* SNP rs389419; the *HLA-DRA* SNP rs9268831; and the *HLA-B* Bw4/Bw6 epitope polymorphism. Supplementary Fig. 1 shows a plot of the regression coefficient for the test locus rs9268831 against number of groups for each bootstrap data set. Notably, although the effect size decreases (that is, the regression coefficient increases) with number of groups, this decrease was very modest, indicating that our effect size estimates are good. We then used the bootstrap samples to calculate 95% confidence intervals for each test locus. Reassuringly, all bootstrap 95% confidence intervals were consistent with the original 95% confidence intervals (rs241448, 95% $CI_{orig} = 0.66$ –1.02 and 95% $CI_{boot} = 0.71$ –1.17; rs389419, 95% $CI_{orig} = 1.13$ –1.59 and 95% $CI_{boot} = 1.17$ –1.84; rs9268831, 95% $CI_{orig} = 0.62$ –0.83 and 95% $CI_{boot} = 0.57$ –0.80; *HLA-B* Bw4/Bw6, 95% $CI_{orig} = 1.02$ –1.45 and 95% $CI_{boot} = 1.003$ –1.45) and do not lead to a different interpretation of results. Hence, we believe that the trees are effective models for the *HLA-DRB1/HLA-DQB1* effects.

Testing for associations at non-class-II loci in families. We specifically wished to test the hypothesis that loci within the MHC were associated with type 1 diabetes independently of the highly associated class II genes *HLA-DRB1* and *HLA-DQB1*. Owing to the complex relationship between these two genes, extensive linkage disequilibrium and epistatic interaction effects², we believed that a joint model was required to explain the observed association. This approach was justified because both loci were necessary to partition the data within recursive partitioning.

Forward stepwise conditional logistic regression was used to test whether any of the 83 loci typed in the MHC had an effect in addition to the HLA class II *DRB1/DQB1* effect²⁸. Only individuals typed at both the class II loci and the test locus were used for the stepwise analysis. The *HLA-DRB1/HLA-DQB1* loci (modelled using the recursive partitioning method described above) were placed in the regression model as confounders and other loci added; whether or not a non-*HLA-DRB1/HLA-DQB1* locus improved on the model was tested by a Wald test where robust variance estimates could be applied, or else by a likelihood ratio test. The non-*HLA-DRB1/HLA-DQB1* loci were modelled as alleles when the multiplicative model was appropriate, and genotypes otherwise.

Testing for non-*HLA-DRB1/HLA-DQB1* loci in the case-control collection. As with the family data set, *HLA-DRB1* and *HLA-DQB1* were grouped by recursive partitioning (detailed above) and placed in the logistic regression model as confounders. A likelihood ratio test was used to test whether other loci added to the regression model. $P < 0.0001$ was considered significant. The analysis was stratified both by broad geographical region³⁰ and by the *HLA-DRB1/HLA-DQB1* groups. The most significant locus was added as alleles or genotypes to the grouped class II loci and other loci added to them to test for additional effects.

Characterizing the cancer genome in lung adenocarcinoma

Barbara A. Weir^{1,2,*}, Michele S. Woo^{1,*}, Gad Getz^{2,*}, Sven Perner^{3,4}, Li Ding⁵, Rameen Beroukhi^{1,2}, William M. Lin^{1,2}, Michael A. Province⁶, Aldi Kraja⁶, Laura A. Johnson³, Kinjal Shah^{1,2}, Mitsuo Sato⁸, Roman K. Thomas^{1,2,9,10}, Justine A. Barletta³, Ingrid B. Borecki⁶, Stephen Broderick^{11,12}, Andrew C. Chang¹⁴, Derek Y. Chiang^{1,2}, Lucian R. Chirieac^{3,16}, Jeonghee Cho¹, Yoshitaka Fujii¹⁸, Adi F. Gazdar⁸, Thomas Giordano¹⁵, Heidi Greulich^{1,2}, Megan Hanna^{1,2}, Bruce E. Johnson¹, Mark G. Kris¹¹, Alex Lash¹¹, Ling Lin⁵, Neal Lindeman^{3,16}, Elaine R. Mardis⁵, John D. McPherson¹⁹, John D. Minna⁸, Margaret B. Morgan¹⁹, Mark Nadel^{1,2}, Mark B. Orringer¹⁴, John R. Osborne⁵, Brad Ozenberger²⁰, Alex H. Ramos^{1,2}, James Robinson², Jack A. Roth²¹, Valerie Rusch¹¹, Hidefumi Sasaki¹⁸, Frances Shepherd²⁵, Carrie Sougnez², Margaret R. Spitz²², Ming-Sound Tsao²⁵, David Twomey², Roel G. W. Verhaak², George M. Weinstock¹⁹, David A. Wheeler¹⁹, Wendy Winckler^{1,2}, Akihiko Yoshizawa¹¹, Soyoung Yu¹, Maureen F. Zakowski¹¹, Qunyuan Zhang⁶, David G. Beer¹⁴, Ignacio I. Wistuba^{23,24}, Mark A. Watson⁷, Levi A. Garraway^{1,2}, Marc Ladanyi^{11,12}, William D. Travis¹¹, William Pao^{11,12}, Mark A. Rubin^{2,3}, Stacey B. Gabriel², Richard A. Gibbs¹⁹, Harold E. Varmus¹³, Richard K. Wilson⁵, Eric S. Lander^{2,17,26} & Matthew Meyerson^{1,2,16}

Somatic alterations in cellular DNA underlie almost all human cancers¹. The prospect of targeted therapies² and the development of high-resolution, genome-wide approaches^{3–8} are now spurring systematic efforts to characterize cancer genomes. Here we report a large-scale project to characterize copy-number alterations in primary lung adenocarcinomas. By analysis of a large collection of tumours ($n = 371$) using dense single nucleotide polymorphism arrays, we identify a total of 57 significantly recurrent events. We find that 26 of 39 autosomal chromosome arms show consistent large-scale copy-number gain or loss, of which only a handful have been linked to a specific gene. We also identify 31 recurrent focal events, including 24 amplifications and 7 homozygous deletions. Only six of these focal events are currently associated with known mutations in lung carcinomas. The most common event, amplification of chromosome 14q13.3, is found in ~12% of samples. On the basis of genomic and functional analyses, we identify *NKX2-1* (NK2 homeobox 1, also called *TTF1*), which lies in the minimal 14q13.3 amplification interval and encodes a lineage-specific transcription factor, as a novel candidate proto-oncogene involved in a significant fraction of lung adenocarcinomas. More generally, our results indicate that many of the genes that are involved in lung adenocarcinoma remain to be discovered.

A collection of 528 snap-frozen lung adenocarcinoma resection specimens, with at least 70% estimated tumour content, was selected by a panel of thoracic pathologists (Supplementary Table 1); samples were anonymized to protect patient privacy. Tumour and normal

DNAs were hybridized to Affymetrix 250K Sty single nucleotide polymorphism (SNP) arrays. Genomic copy number for each of over 238,000 probe sets was determined by calculating the intensity ratio between the tumour DNA and the average of a set of normal DNAs^{9,10}. Segmented copy numbers for each tumour were inferred with the GLAD (gain and loss analysis of DNA) algorithm¹¹ and normalized to a median of two copies. Each copy number profile was then subjected to quality control, resulting in 371 high-quality samples used for further analysis, of which 242 had matched normal samples (Methods).

To identify regions of copy-number alteration, we applied GISTIC (genomic identification of significant targets in cancer)¹², a statistical method that calculates a score that is based on both the amplitude and frequency of copy-number changes at each position in the genome, using permutation testing to determine significance (Methods).

GISTIC identified 26 large-scale events and 31 focal events, reported below. Although the overall pattern is broadly consistent with the literature on lung cancer^{8,13–15}, our sample size and resolution provide more power to accurately identify and localize both large-scale and focal chromosomal alterations. With respect to large-scale events, no single previous study has identified more than 5 of the gains or 11 of the losses^{13,14} (Supplementary Table 2). With respect to focal events, three recent studies^{8,14,15} report a total of ~200 events, including 23 of the 31 recurrent focal events observed in our study. The overlap among these three studies is limited to only four events (amplification of *EGFR*, *CCNE1*, *MDM2* and 8p11, all seen

¹Department of Medical Oncology and Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ²Cancer Program, Genetic Analysis Platform, and Genome Biology Program, Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. ³Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. ⁴Institute of Pathology, University of Ulm, Ulm 89081, Germany. ⁵Genome Sequencing Center, ⁶Division of Statistical Genomics and ⁷Department of Pathology and Immunology, Washington University in Saint Louis, Saint Louis, Missouri 63130, USA. ⁸University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ⁹Max Planck Institute for Neurological Research with Klaus-Joachim Zülch Laboratories of the Max-Planck Society and the Medical Faculty of the University of Cologne, Cologne 50931, Germany. ¹⁰Center for Integrated Oncology and Department I for Internal Medicine, University of Cologne, Cologne 50931, Germany. ¹¹Departments of Medicine, Surgery, Pathology, and Computational Biology, ¹²Human Oncology and Pathogenesis Program, ¹³Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ¹⁴Section of Thoracic Surgery, Department of Surgery and ¹⁵Department of Pathology, University of Michigan, Ann Arbor, Michigan 48109, USA. ¹⁶Department of Pathology and ¹⁷Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹⁸Department of Surgery, Nagoya City University Medical School, Nagoya 467-8602, Japan. ¹⁹Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA. ²⁰National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. ²¹Department of Thoracic and Cardiovascular Surgery, ²²Department of Epidemiology, ²³Department of Pathology and ²⁴Department of Thoracic/Head and Neck Medical Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030, USA. ²⁵University Health Network and Princess Margaret Hospital, Toronto M5G 2C4, Canada. ²⁶Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA.

*These authors contributed equally to this work.

here; Supplementary Table 3 and Supplementary Results). A genome-wide view of segmented copy number reveals that most chromosomal arms undergo either amplification or deletion across a large proportion of the samples (Fig. 1a). The distinctive pattern of amplification and loss is also apparent when the median copy number for each chromosome arm is plotted (Supplementary Fig. 1 and Supplementary Table 4). In total, GISTIC identifies 26 large segments (at least half of a chromosome arm), 10 with significant gains and 16 with significant losses (Fig. 1b and Supplementary Table 5).

Visual inspection reveals that similar chromosomal patterns of copy number loss and gain across the genome are found in almost all samples, but that the samples show substantial differences in the amplitude of copy-number variation (Fig. 1a). When the samples are

partitioned into tertiles on the basis of overall variation in copy-number amplitude, each shows similar regions of amplification and loss across the genome. The attenuation seen in many samples is consistent with admixture with euploid non-tumour DNA, which we estimate at 50%, 65% and 78% respectively in the three tertiles (Supplementary Results and Supplementary Fig. 2). The significant non-tumour admixture in these tumour samples also makes it difficult to assess genome-wide loss of heterozygosity (LOH). Because normal DNA admixture limits sensitivity, we report LOH only in the top tertile; we see both LOH associated with copy-number loss and copy-neutral LOH (chromosomes 17p and 19p) (Supplementary Results, Supplementary Figs 3 and 4, and Supplementary Table 6).

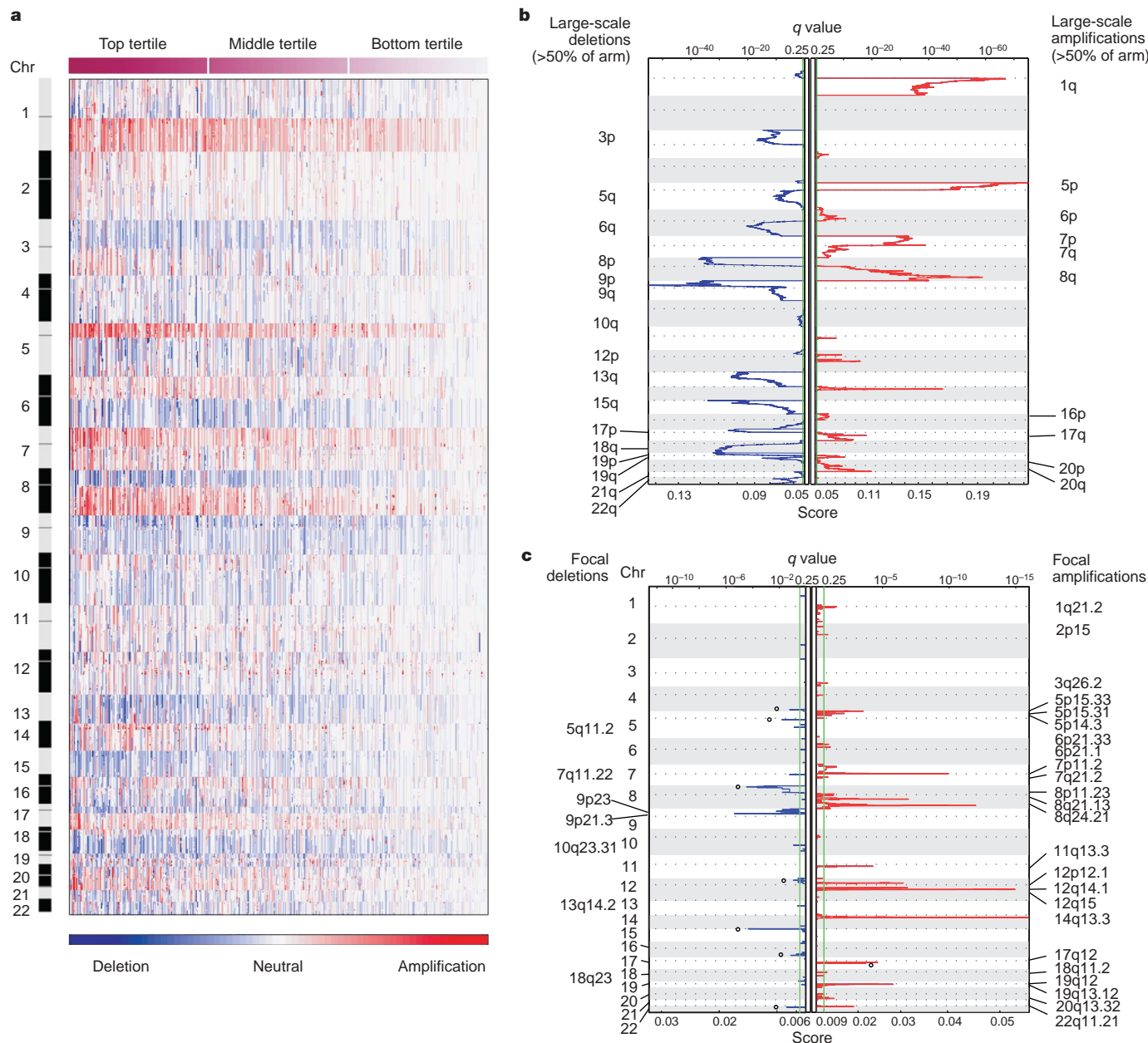


Figure 1 | Large-scale genomic events in lung adenocarcinoma. **a**, Smoothed copy number data for 371 lung adenocarcinoma samples (columns; ordered by degree of interchromosomal variation and divided into top, middle and bottom tertiles) is shown by genomic location (rows). The colour scale ranges from blue (deletion) through white (neutral; two copies in diploid specimens) to red (amplification). **b**, **c**, False-discovery rates (q values; green line is 0.25 cut-off for significance) and scores for each

alteration (x axis) are plotted at each genome position (y axis); dotted lines indicate the centromeres. Amplifications (red lines) and deletions (blue lines) are shown for large-scale events (**b**; $\geq 50\%$ of a chromosome arm; copy number threshold = 2.14 and 1.87) and focal events (**c**; copy number threshold = 3.6 and 1.2). Open circles label known or presumed germline copy-number polymorphisms.

The most common genomic alteration in lung adenocarcinoma is copy-number gain of chromosome 5p, which is found in 60% of total samples and over 80% of the top tertile (Supplementary Table 5). Another 15 large-scale events are seen in at least 33% of all samples and over 40% of the top tertile. Together, the regions of common copy-number gain (~650 megabases (Mb)) and copy-number loss (~1,010 Mb) comprise more than half of the human genome (Supplementary Results and Supplementary Table 5). Despite their high frequency, few of these large-scale events have been clearly related to functional effects on specific genes. Loss of a chromosome arm is likely to act by uncovering an inactivated tumour suppressor gene, yet such mutations have been well-established in lung adenocarcinoma in only three of the sixteen deleted chromosome arms (*CDKN2A* on 9p, *TP53* on 17p and *STK11* on 19p)^{16–18}. We tested for correlations between the large-scale lesions and clinical parameters, but none was significant after correction for multiple hypothesis testing (Supplementary Results and Supplementary Table 7).

Focal deletions may help pinpoint tumour suppressor genes, particularly on chromosome arms that show frequent copy-number loss. At a threshold set to detect homozygous deletions in the presence of stromal contamination, GISTIC analysis identified seven focal candidate regions (Fig. 1c and Supplementary Table 8). The most significant focal deletions, detected in 3% of all samples and 6.5% of the top tertile, encompass *CDKN2A/CDKN2B*, two well-documented tumour suppressor genes on chromosome 9p21 (Fig. 1c, Table 1 and Supplementary Table 8). The protein products of *CDKN2A* and *CDKN2B* inhibit two cyclin-dependent kinases, Cdk4 and Cdk6, the genes of which both reside in frequently amplified regions (see below). Two other deleted regions also encompass known tumour suppressor genes, *PTEN* and *RBI* (Supplementary Table 8).

Three additional deletion regions each localize to a single gene. Deletions of the 5' untranslated region of *PTPRD*, encoding a tyrosine phosphatase, occur in 4% of the top tertile. Although *PTPRD* deletions have been reported in lung adenocarcinoma cell lines^{8,19,20}, this is the first observation in primary human lung adenocarcinomas. Homozygous deletions of *PDE4D* occur in 1.6% of the top tertile and typically remove several hundred kilobases and affect multiple exons (Supplementary Fig. 5). These deletions may be significant for lung

biology because *PDE4D* encodes the major phosphodiesterase responsible for degrading cyclic AMP in airway epithelial cells²¹. Another single-gene deletion occurs within *AUTS2*, a gene of unknown function in chromosome 7q11.22 (Table 1 and Supplementary Table 8). We cannot exclude the possibility that some recurrent copy-number losses are due to genomic fragility unrelated to carcinogenesis; the presence of point mutations would provide additional support for a role in cancer.

We therefore sequenced all exons of *AUTS2*, *PDE4D* and *PTPRD*, as each of these genes showed single-gene deletions but no mutations have been reported in primary tumours. Although we did not detect somatic mutations in *AUTS2* or *PDE4D*, we identified validated somatic *PTPRD* mutations in 11 of 188 samples sequenced. Notably, three of the mutations encode predicted inactivating changes in the tyrosine phosphatase domain (Supplementary Table 9 and Supplementary Fig. 6). These results implicate *PTPRD* as a probable cancer-associated gene, although further studies are needed to establish a causative role in cancer via gain or loss of function.

We focused above on homozygous deletions, but note that this approach will miss important genes. Notably, the *TP53* locus is known to be mutated in ~50% of lung adenocarcinomas but shows no homozygous deletions in our data.

We next focused on focal amplification events, for which it may be easier to pinpoint target genes. At a threshold designed to identify high-copy amplification, the GISTIC analysis identified 24 recurrent genomic segments with maximum copy number ranging from about 4- to 16-fold (Fig. 1c, Table 1 and Supplementary Table 10). The amplification events are seen in 1–7% of all samples (1–12% in the top tertile). Each of these events is seen in at least two samples and all but eight are seen in at least five samples. In the 13 most significant amplifications ($q < 0.01$), the regions can be localized to relatively small genomic segments containing 15 or fewer genes. Although 14 of the 24 regions of recurrent amplification contain a known proto-oncogene (Supplementary Table 10), only three of these genes (*EGFR*, *KRAS* and *ERBB2*) have been previously reported to be mutated in lung adenocarcinoma (Supplementary Results). The remaining 11 genes are clear targets for re-sequencing in lung tumours.

Table 1 | Top focal regions of amplification and deletion

Cytoband*	q value	Peak region (Mb)*	Max/Min inferred copy no.	Number of genes*†	Known proto-oncogene/tumour suppressor gene in region*‡	New candidate(s)
Amplifications						
14q13.3	2.26×10^{-29}	35.61–36.09	13.7	2	–	<i>NKX2-1</i> , <i>MBIP</i>
12q15	1.78×10^{-15}	67.48–68.02	9.7	3	<i>MDM2</i>	–
8q24.21	9.06×10^{-13}	129.18–129.34	10.3	0	<i>MYC</i> §	–
7p11.2	9.97×10^{-11}	54.65–55.52	8.7	3	<i>EGFR</i>	–
8q21.13	1.13×10^{-7}	80.66–82.55	10.4	8	–	–
12q14.1	1.29×10^{-7}	56.23–56.54	10.4	15	<i>CDK4</i>	–
12p12.1	2.83×10^{-7}	24.99–25.78	10.4	6	<i>KRAS</i>	–
19q12	1.60×10^{-6}	34.79–35.42	6.7	5	<i>CCNE1</i>	–
17q12	2.34×10^{-5}	34.80–35.18	16.1	12	<i>ERBB2</i>	–
11q13.3	5.17×10^{-5}	68.52–69.36	6.5	9	<i>CCND1</i>	–
5p15.33	0.000279	0.75–1.62	4.2	10	<i>TERT</i>	–
22q11.21	0.001461	19.06–20.13	6.6	15	–	–
5p15.31	0.007472	8.88–10.51	5.6	7	–	–
1q21.2	0.028766	143.48–149.41	4.6	86	<i>ARNT</i>	–
20q13.32	0.0445	55.52–56.30	4.4	6	–	–
5p14.3	0.064673	19.72–23.09	3.8	2	–	–
6p21.1	0.078061	43.76–44.12	7.7	2	–	<i>VEGFA</i>
Deletions						
9p21.3	3.35×10^{-13}	21.80–22.19	0.7	3	<i>CDKN2A/CDKN2B</i>	–
9p23	0.001149	9.41–10.40	0.4	1	–	<i>PTPRD</i>
5q11.2	0.005202	58.40–59.06	0.6	1	–	<i>PDE4D</i>
7q11.22	0.025552	69.50–69.62	0.7	1	–	<i>AUTS2</i>
10q23.31	0.065006	89.67–89.95	0.5	1	<i>PTEN</i>	–

* Based on hg17 human genome assembly.

† RefSeq genes only.

‡ Known tumour suppressor genes and proto-oncogenes defined as found in either COSMIC³⁰, CGP Census³¹ or other evidence; if there is more than one known proto-oncogene in the region, only one is listed (priority for listing is, in order: known lung adenocarcinoma mutation; known lung cancer mutation; other known mutation (by COSMIC frequency); listing in CGP Census).

§ *MYC* is near, but not within, the peak region.

|| Single gene deletions previously seen, this study provides new mutations as well.

Our data localize the amplification peak on chromosome 5p to the telomerase catalytic subunit gene, *TERT*. Although broad amplification of chromosome 5p has been described in non-small-cell lung cancer (NSCLC)^{13,22,23}, the target of 5p amplification has not been determined. In our data set, eight tumours with amplicons in chromosome 5p15 delineate a region containing ten genes, including *TERT* (Table 1 and Supplementary Table 10), suggesting that *TERT* may be the target of the amplification and thereby contributes to cellular immortalization.

Chromosome 6p21.1 shows focal amplification in four samples in a region containing two genes, one of which (*VEGFA*) encodes vascular endothelial growth factor (Table 1 and Supplementary Table 10). This amplification suggests a possible mechanism for increased angiogenesis and for the reported response to angiogenic inhibitors such as the anti-VEGF antibody bevacizumab in lung adenocarcinoma^{24,25}. Similarly, amplification of regions including several cell cycle genes such as *CDK4*, *CDK6* and *CCND1* suggests an important role for these genes (Table 1 and Supplementary Table 10).

Notably, the most common focal amplification does not include any known proto-oncogenes: chromosome 14q13.3 is amplified in 6% of the samples overall and 12% of the samples in the top tertile (Fig. 1c, Table 1 and Supplementary Table 10; $q < 10^{-28}$). Although previous studies have reported amplification of 14q13 in lung cancer cell lines¹⁴ and the region is mentioned in studies of primary lung tumours^{8,15}, the target gene in this region had not been identified. With our large sample size, we are able to narrow the critical region to a 480-kilobase interval containing only two known genes, *MBIP* and *NKX2-1* (Fig. 2a, b, Table 1 and Supplementary Table 10). Data for a single tumour with a small region of high-level amplification, comprising *MBIP* and *NKX2-1*, exclude the neighbouring gene, *NKX2-8* (Fig. 2c).

We confirmed the amplification of the region by fluorescence *in situ* hybridization (FISH) and quantitative polymerase chain reaction (qPCR; data not shown). FISH analysis was performed with a bacterial artificial chromosome (BAC) probe containing *NKX2-1* and *NKX2-8* (Fig. 2c) on an independent set of 330 lung adenocarcinoma samples from tissue microarrays. High-level amplification of the chromosome 14q13.3 region was seen in 12% (40 out of 330) of these lung tumours. The FISH studies revealed amplification up to an estimated 100-fold (Fig. 2d and Supplementary Fig. 7); the lower amplification estimated on the SNP arrays (up to 14-fold) probably reflects signal saturation, stromal admixture and tumour heterogeneity. No significant difference in patient survival after surgical resection and long-term follow-up was observed between tumours with amplified or non-amplified *NKX2-1* (Supplementary Fig. 8 and Supplementary Table 11). Exon-based sequencing in 384 lung adenocarcinoma DNA samples showed no somatic mutations in either *NKX2-1* or *MBIP* (Supplementary Results), indicating that any oncogenic function might be exerted by the wild-type gene.

We used RNA interference (RNAi) to test the roles of both *MBIP* and *NKX2-1* with respect to cell survival and oncogenic properties. Expression of two different short hairpin RNAs (shRNAs) targeting *NKX2-1* significantly reduced the levels of *NKX2-1* protein in NCI-H2009 cells (Fig. 3a) and NCI-H661 cells (data not shown)—NSCLC lines that carry 14q amplifications¹⁴. No *NKX2-1* protein was detected in A549 cells that lack 14q amplification (Fig. 3a).

RNAi-mediated inhibition of *NKX2-1* expression substantially decreased the ability of NCI-H2009 cells to grow in an anchorage-independent manner as measured by colony formation in soft agar (Fig. 3b), which may be due, in part, to a loss of cell viability. NCI-H661 cell viability was also impaired by *NKX2-1* RNAi (Supplementary Fig. 9). *NKX2-1* knockdown leads to a decrease in colony formation in lung adenocarcinoma lines (NCI-H1975 and HCC1171) that lack chromosome 14q13 amplification but express *NKX2-1* (Supplementary Fig. 10), but has no effect on either soft agar colony formation or cell viability in A549 cells, which express little or no *NKX2-1* protein (Fig. 3a, c). In contrast to the results for *NKX2-1*,

RNAi-based *MBIP* knockdown neither decreased colony formation in NCI-H2009 cells (Fig. 3d, e) or in NCI-H661 cells (Supplementary Fig. 11a, b), nor reduced cell viability (Supplementary Fig. 11c, d). It thus seems that *NKX2-1*, but not *MBIP*, is essential for the survival and tumorigenic properties of lung adenocarcinoma cell lines that express *NKX2-1*.

Systematic understanding of the molecular basis of a particular type of cancer will require at least three steps: comprehensive characterization of recurrent genomic aberrations (including copy-number changes, nucleotide sequence changes, chromosomal rearrangements and epigenetic alterations); elucidation of their biological role in cancer pathogenesis; and evaluation of their utility for diagnostics, prognostics and therapeutics. This study represents a step towards comprehensive genomic characterization of one of the

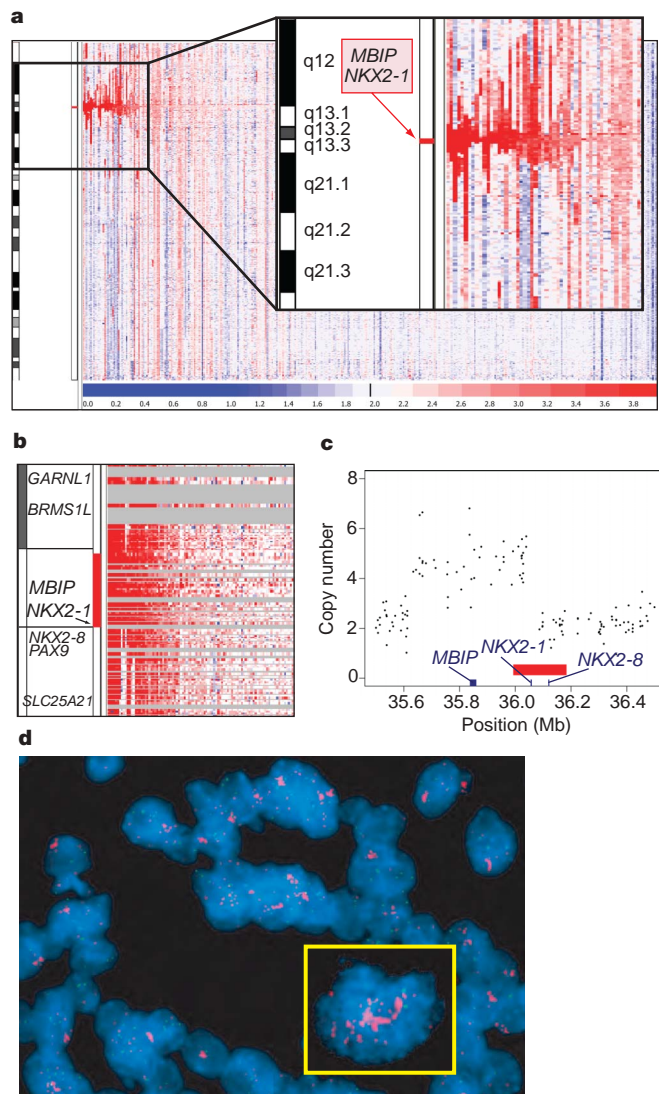


Figure 2 | High-prevalence amplification of the *MBIP/NKX2-1* locus on chromosome 14q. **a**, Copy number on chromosome 14q is shown for 371 lung adenocarcinomas (columns; ordered by amplification) from centromere (top) to telomere (bottom). Colour scale as in Fig. 1. **b**, Magnified view of the amplified region from **a**; grey bars represent the absence of SNPs on the array. **c**, Raw copy number data (y axis) for one sample defining the minimally amplified region are plotted according to chromosome 14 position (x axis; scale in megabases). Genomic positions of *MBIP*, *NKX2-1*, *NKX2-8* and the BAC used for FISH (red bar) are shown along the x axis. **d**, FISH for *NKX2-1* (red) and a chromosome 14 reference probe (green) on a lung adenocarcinoma specimen with high-level amplification of the *NKX2-1* probe. Nuclei are stained with 4,6-diamidino-2-phenylindole (DAPI; blue). The yellow box shows a single nucleus.

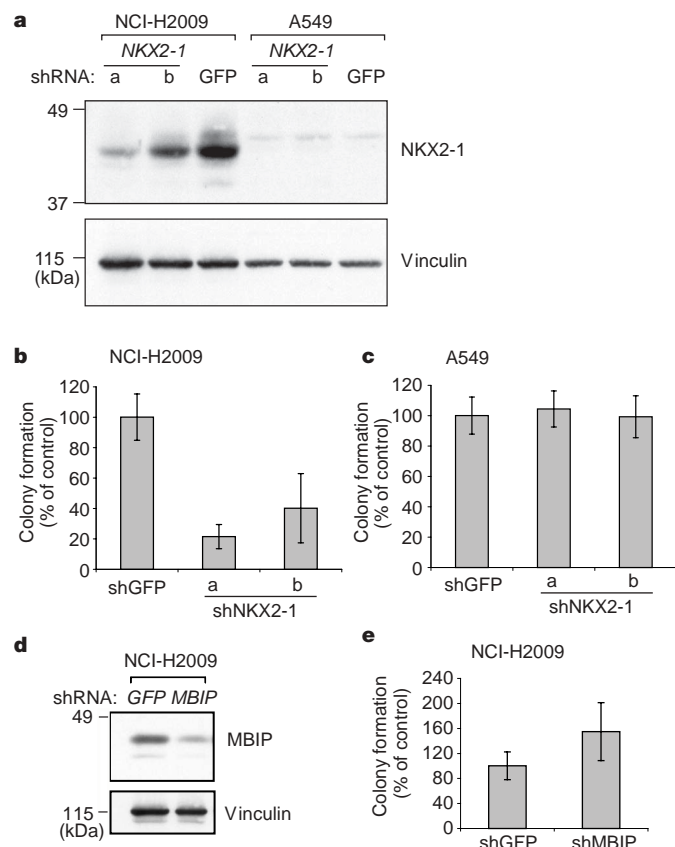


Figure 3 | *NKX2-1* RNAi leads to reduced anchorage-independent growth and viability of NCI-H2009 cells but not A549 cells. **a**, Anti-*NKX2-1* and control anti-vinculin immunoblots of lysates from NCI-H2009 and A549 cells expressing shRNA against *NKX2-1* (shNKX2-1a and shNKX2-1b) or GFP (shGFP) as control. **b**, Soft agar colony formation by NCI-H2009 cells is shown relative to the shGFP control as a mean percentage (\pm standard deviation in triplicate samples; $P = 5.8 \times 10^{-6}$ when comparing shGFP to shNKX2-1a and $P = 5.1 \times 10^{-4}$ when comparing shGFP to shNKX2-1b). **c**, Colony formation assays as in **b** for A549 cells ($P > 0.5$). **d**, Anti-MBIP and anti-vinculin immunoblots of lysates from shRNA-expressing NCI-H2009 cells. **e**, Colony formation of shMBIP NCI-H2009 cells relative to that of control shGFP cells ($P = 0.0344$).

most common cancers, lung adenocarcinoma. We define two main types of recurrent events in this disease: frequent, large-scale events and rare, focal events. Further efforts to identify the target genes of the frequent, large-scale events will probably involve systematic screens to produce orthogonal data sets (mutational, epigenetic, expression and loss-of-function phenotypes).

Strikingly, the single most common focal event in lung adenocarcinoma (amplification of 14q13.3) was not previously associated with a specific gene. We show here that the target gene is *NKX2-1*, a transcription factor that has an essential role in the formation of type II pneumocytes, the cell type that lines the alveoli of the lung^{26,27}. *Nkx2-1* knockout mice fail to develop normal type II pneumocytes or alveoli and die of respiratory insufficiency at birth²⁸, which highlights the importance of *NKX2-1* in lung development. *NKX2-1* shows hallmarks of a novel lineage-survival oncogene, similar to the *MITF* gene in melanoma⁷. The lineage-restricted amplification of such genes contrasts with the more ubiquitous amplifications seen for genes in cell cycle (for example, *CDK4*, *CDK6*, *CCND1*, *CCNE1*) and signal transduction (for example, *EGFR*, *ERBB2*, *KRAS*) pathways.

More generally, our results, together with other recent studies²⁹, illustrate the power of systematic copy-number analysis with SNP arrays. They make clear that many important cancer-related genes remain to be discovered and can be revealed by systematic genomic study.

METHODS SUMMARY

DNA specimens were labelled and hybridized to the Affymetrix 250K Sty I array to obtain signal intensities and genotype calls. Loci identified by GISTIC analysis were further characterized by sequencing, genotype validation, tissue microarray FISH and functional studies. RNAi was performed by stable expression of shRNA lentiviral vectors targeting *NKX2-1*, *MBIP* or *GFP* in lung cancer cell lines, which were then used in soft agar and cell proliferation assays. Raw data and related files are available at <http://www.broad.mit.edu/tsp>.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 April; accepted 10 October 2007.

Published online 4 November 2007.

- Weir, B., Zhao, X. & Meyerson, M. Somatic alterations in the human cancer genome. *Cancer Cell* **6**, 433–438 (2004).
- Sawyers, C. Targeted cancer therapy. *Nature* **432**, 294–297 (2004).
- Pinkel, D. et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nature Genet.* **20**, 207–211 (1998).
- Pollack, J. R. et al. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genet.* **23**, 41–46 (1999).
- Bignell, G. R. et al. High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res.* **14**, 287–295 (2004).
- Zhao, X. et al. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res.* **64**, 3060–3071 (2004).
- Garraway, L. A. et al. Integrative genomic analyses identify *MITF* as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436**, 117–122 (2005).
- Zhao, X. et al. Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. *Cancer Res.* **65**, 5561–5570 (2005).
- Li, C. & Wong, W. H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl Acad. Sci. USA* **98**, 31–36 (2001).
- Li, C. & Wong, W. H. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol.* **2**, RESEARCH0032 (2001).
- Hu, P., Stransky, N., Thiery, J. P., Radvanyi, F. & Barillot, E. Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics* **20**, 3413–3422 (2004).
- Beroukhi, R. et al. Assessing the significance of chromosomal aberrations in cancer: Methodology and application to glioma. *Proc. Natl Acad. Sci. USA*. (in the press).
- Balsara, B. R. & Testa, J. R. Chromosomal imbalances in human lung cancer. *Oncogene* **21**, 6877–6883 (2002).
- Garnis, C. et al. High resolution analysis of non-small cell lung cancer cell lines by whole genome tiling path array CGH. *Int. J. Cancer* **118**, 1556–1564 (2006).
- Tonon, G. et al. High-resolution genomic profiles of human lung cancer. *Proc. Natl Acad. Sci. USA* **102**, 9625–9630 (2005).
- Hayashi, N., Sugimoto, Y., Tsuchiya, E., Ogawa, M. & Nakamura, Y. Somatic mutations of the *MTS* (multiple tumor suppressor) 1/*CDK4* (cyclin-dependent kinase-4 inhibitor) gene in human primary non-small cell lung carcinomas. *Biochem. Biophys. Res. Commun.* **202**, 1426–1430 (1994).
- Sanchez-Cespedes, M. et al. Inactivation of *LKB1/STK11* is a common event in adenocarcinomas of the lung. *Cancer Res.* **62**, 3659–3662 (2002).
- Takahashi, T. et al. p53: a frequent target for genetic abnormalities in lung cancer. *Science* **246**, 491–494 (1989).
- Sato, M. et al. Identification of chromosome arm 9p as the most frequent target of homozygous deletions in lung cancer. *Genes Chromosomes Cancer* **44**, 405–414 (2005).
- Cox, C. et al. A survey of homozygous deletions in human cancer genomes. *Proc. Natl Acad. Sci. USA* **102**, 4542–4547 (2005).
- Barnes, A. P. et al. Phosphodiesterase 4D forms a cAMP diffusion barrier at the apical membrane of the airway epithelium. *J. Biol. Chem.* **280**, 7997–8003 (2005).
- Zhu, C. Q. et al. Amplification of telomerase (*hTERT*) gene is a poor prognostic marker in non-small-cell lung cancer. *Br. J. Cancer* **94**, 1452–1459 (2006).
- Zhang, A. et al. Frequent amplification of the telomerase reverse transcriptase gene in human tumors. *Cancer Res.* **60**, 6230–6235 (2000).
- Johnson, D. H. et al. Randomized phase II trial comparing bevacizumab plus carboplatin and paclitaxel with carboplatin and paclitaxel alone in previously untreated locally advanced or metastatic non-small-cell lung cancer. *J. Clin. Oncol.* **22**, 2184–2191 (2004).
- Sandler, A. et al. Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. *N. Engl. J. Med.* **355**, 2542–2550 (2006).
- Bingle, C. D. Thyroid transcription factor-1. *Int. J. Biochem. Cell Biol.* **29**, 1471–1473 (1997).
- Ikeda, K. et al. Gene structure and expression of human thyroid transcription factor-1 in respiratory epithelial cells. *J. Biol. Chem.* **270**, 8108–8114 (1995).

28. Yuan, B. *et al.* Inhibition of distal lung morphogenesis in *Nkx2.1*^{-/-} embryos. *Dev. Dyn.* **217**, 180–190 (2000).
29. Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758–764 (2007).
30. Bamford, S. *et al.* The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br. J. Cancer* **91**, 355–358 (2004).
31. Futreal, P. A. *et al.* A census of human cancer genes. *Nature Rev. Cancer* **4**, 177–183 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from the US National Cancer Institute (B.A.W., M.S.W., M.R.S., M.M., I.I.W., A.F.G., J.A.R., M.S., J.D.M.), the US National Human Genome Research Institute (R.A.G., R.K.W., E.S.L.), the Canadian Cancer Society/National Cancer Institute (M.S.T.), the American Lung Association (M.M.), Joan's Legacy Foundation (M.M.), the American Cancer Society (M.M.), the International Association for the Study of Lung Cancer (R.K.T.), the US Department of Defense (R.B., I.I.W., J.D.M.) and the Carmel Hill Fund (W.P., M.G.K., H.E.V.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.M. (matthew_meyerson@dfci.harvard.edu).

METHODS

Primary lung specimens. A total of 575 DNA specimens were obtained from primary lung tumours (all of them with the original diagnosis of lung adenocarcinoma, 528 of which were confirmed to be lung adenocarcinomas), 439 matched normal samples and 53 additional normal specimens. These DNAs were labelled and hybridized to SNP arrays (see below) without previous whole-genome amplification. Each of the selected tumour samples was determined to have greater than 70% tumour percentage by pathology review.

Of the 575 selected tumours, 384 anonymous lung tumour and matched normal DNAs for the Tumour Sequencing Project (TSP) were collected from five sites: Memorial-Sloan Kettering Cancer Center (102 tumours and paired normal samples), University of Michigan (101 tumours and paired normal samples), MD Anderson Cancer Center (29 tumours and paired normal samples), Washington University (84 tumours and paired normal samples) and Dana-Farber Cancer Institute/The Broad Institute (68 tumours and paired normal samples). Additional anonymous lung adenocarcinoma samples or DNAs were collected from the Brigham and Women's Hospital tissue bank (19 tumours and 18 paired normal samples), H. Sasaki at the Nagoya City University Medical School (112 tumours and 37 paired normal samples) and from the University Health Network in Toronto (60 tumour samples). In addition to the matched normal samples, 53 unmatched normal tissue or blood samples were used for SNP array normalization purposes (sources include J. Llovet, S. Pomeroy, S. Singer, the Genomics Collaborative, Inc., Massachusetts General Hospital and R. Beroukhim). All tumour samples were surgically dissected and frozen at -80°C until use.

SNP array experiments. For each sample, SNPs were genotyped with the Sty I chip of the 500K Human Mapping Array set (Affymetrix Inc.). Array experiments were performed according to manufacturer's directions. In brief, for each sample, 250 ng of genomic DNA was digested with the StyI restriction enzyme (New England Biolabs). The digested DNA was then ligated to an adaptor with T4 ligase (New England Biolabs) and PCR-amplified using an Applied Biosystems 9700 Thermal Cycler I and Titanium Taq (Clontech) to achieve a size range of 200–1,100 bp. Amplified DNA was then pooled, concentrated and put through a clean-up set. The product was then fragmented using DNaseI (Affymetrix Inc.) and subsequently labelled, denatured and hybridized to arrays. Hybridized arrays were scanned using the GeneChip Scanner 3000 7G (Affymetrix Inc.). Batches of 96 samples were processed as a single plate using a Biomek FX robot with dual 96 and span-8 heads (Beckman Coulter) and a GeneChip Fluidics Station FS450 (Affymetrix Inc.). Samples and plates were tracked using ABGene 2D barcode rack and single tube readers (ABGene). Tumour and paired normal sample (where applicable) were always placed in adjacent wells on the same plate to minimize experimental differences. Raw data (.CEL and .txt files) are available at <http://www.broad.mit.edu/tsp>.

Primary SNP array data analysis. SNP arrays were processed as a plate of 96 samples using the GenePattern software package³², with modules based on dChipSNP algorithms^{9,10}. GenePattern modules are available at <http://www.broad.mit.edu/cancer/software/genepattern/>. Intensity (.CEL) files were normalized and modelled using the PM-MM difference modelling method⁹ with the SNPfileCreator module. Array normalization, similar to quantile normalization, was performed³³; 6,000 matching quantiles from the probe density distributions of two arrays were used to fit a running median normalization curve for normalization of each array to a common baseline array¹⁰.

Array quality control analysis. Further analysis was performed on arrays that met certain quality control criteria. As a first step, non-adenocarcinoma samples ($n = 47$) from the TSP set of 384 tumours were removed from further analysis (leaving 528 adenocarcinomas). Technical failure criteria (removing 33 tumours) included a requirement for correct tumour/normal matching, genotyping call rates (% of SNPs that a genotype call can be inferred for) greater than 85% and a score measuring copy-number variation between neighbouring SNPs of less than 0.5. The measure of local SNP copy number variation is calculated by the formula: variation score = $\text{mean}[(\log(\text{RC}_i) - \log(\text{RC}_{i+1}))^2 + (\log(\text{RC}_i) - \log(\text{RC}_{i-1}))^2]$, where RC_i is the raw copy number at SNP i and the mean is taken over all SNPs. Criteria also included a requirement that after taking the log₂ ratio and performing segmentation by GLAD¹¹, the number of times the smoothed copy number crossed ± 0.1 on the log scale in the genome of tumour samples was < 100 (removing 73 tumours). The same test was used to exclude normal samples, with the number of times the smoothed copy number crossed ± 0.1 decreased to < 45 (removing 50 normal samples). A histogram quality control step, as part of the GISTIC procedure, then removed tumours ($n = 51$) with high degrees of non-tumour DNA contamination by looking for samples with only one peak of copy number across its whole genome. This histogram quality control step also removed normals ($n = 20$) with tumour DNA contamination by looking for samples with greater than one peak of copy number across its whole genome.

GISTIC analysis. GISTIC analysis¹² was performed on arrays that met certain quality control criteria. Raw intensity value files from the GenePattern SNPfileCreator module were used as input into the GISTIC algorithm. In brief, batch correction, data normalization, copy-number determination using either the paired normal sample or the average of the five closest normal samples and copy number segmentation was performed. Data-set-specific copy number polymorphisms were identified by running GISTIC on the set of normal samples alone; the regions identified from this analysis were then also removed from the subsequent analysis of tumours. GISTIC then assigns G^{AMP} and G^{DEL} scores to each locus, respectively representing the frequency of amplifications (deletions) seen at that locus, multiplied by the average increase (decrease) in the log₂ ratio in the amplified (deleted) samples. The score (G) is based on the average amplitude (a) of the lesion type (amplification or deletion) and its frequency (f) in the data set according to the formula: $G_i^{(\text{lesion type})} = f_i^{(\text{lesion type})} a_i^{(\text{lesion type})}$. The significance of each score is determined by comparison to similar scores obtained after permuting the data within each sample. The resulting q -value is an upper bound for the expected fraction of false positives among all regions with a particular q -value or less. GISTIC also implements a peel-off step, which identifies additional secondary peaks within a region.

GISTIC analysis was performed essentially the same as is described in a future publication¹², with the following exceptions. Copy number determination was performed for each tumour using its matched normal sample when available and of good quality ($n = 242$). For all others, the average of the five closest normal samples was used ($n = 129$). Copy number segmentation was performed using the GLAD algorithm with parameter $d = 10$. GLAD segments less than eight SNPs in length were also removed.

Regions identified by GISTIC were also compared to known copy-number polymorphisms³⁴ and were manually reviewed for the presence of the alteration in the paired normal sample. Focal deletion regions with events that occurred in tumour samples that did not have paired normal samples were considered presumed polymorphisms and also removed from the list. Secondary peaks and known and presumed germline copy number polymorphisms are listed in Supplementary Tables 12 and 13.

GISTIC analysis of large-scale regions. Significant broad regions of amplification and deletion were identified by applying GISTIC with the default thresholds of 2.14/1.87 (log₂ ratio of ± 0.1). Regions identified by GISTIC that were greater than 50% of a chromosome arm were considered large-scale. Region frequencies were calculated by determining the number of samples that had a median log₂ ratio greater/less than the threshold (± 0.1), for those SNPs within the region.

GISTIC analysis of focal regions. Significant focal regions of amplification and deletion were identified by applying GISTIC with a threshold of 3.6/1.2 (log₂ ratio of 0.848/ -0.737).

Data visualization. Normalized raw copy number from GISTIC analysis was used as input for visualization in the GenePattern SNPviewer (<http://www.broad.mit.edu/cancer/software/genepattern/>)³². Mapping information for SNP, Refgene and cytoband locations are based on Affymetrix annotations and the hg17 build of the human genome sequence from the University of California, Santa Cruz (<http://genome.ucsc.edu>).

Chromosome arm analysis. After segmentation by GLAD, the median of each chromosome arm for each sample was calculated. Amplification or deletion of an arm across the data set was tested for significance by a two-sided binomial test, after removing log₂ copy number ratios between ± 0.1 . P values were false-discovery rate (FDR) corrected to give a FDR q value; significance is set to a q value of 0.01. The standard deviation of the median copy number of significant arms was then used to sort samples into three groups. Higher standard deviation implies higher interchromosomal variation, which correlates with less stromal contamination. Frequencies were then calculated for the total set and for only the top one-third least stromally contaminated samples to give a better idea of true frequencies in the context of attenuated signal owing to stromal contamination.

Comparison between tertiles. A similar chromosome arm analysis was performed independently on the three sample groups, separated according to the standard deviations of their median arm log₂ copy number ratios. Amplification or deletion of an arm across the data set was tested for significance by a two-sided binomial test, after removing values between ± 0.0125 . P values were FDR corrected to give a FDR q value; significance is set to a q value of 0.01.

Estimation of stromal contamination. To attempt to estimate stromal contamination, we calculated the allele-specific copy numbers by taking all informative SNPs in each of the 237 tumours that have a paired normal (removing five bad pairs) and dividing the allele-specific signal from the tumour by that of the normal. Then for each SNP we found M , the minimum between the copy numbers of the A and B alleles. In regions in which one allele has zero copies (for example, one copy loss in diploid cells) M represents the stromal contamination level (as the stroma has one copy of each allele). We calculated the median value

of *M* across each of the chromosome arms and then estimated the stromal contamination by taking their minimum.

LOH analysis. Inferred LOH calls using an HMM algorithm for 242 tumour/normal sample pairs were generated using dChipSNP³⁵. Default parameters were used, except the genotyping error rate was set to 0.2. Five bad-quality sample pairs were removed before visualization and GISTIC analysis. GISTIC analysis of LOH calls and copy loss for 237 samples were performed as described¹².

Correlation analysis. Associations were tested between each large-scale alteration identified by GISTIC and certain clinical parameters. A Fisher's exact test was used to determine association of large-scale copy-number lesions with the binary clinical parameters (gender and smoking status). A chi-squared test was used to determine whether each large-scale copy number alteration was independent of each non-binary clinical parameter (age range, differentiation, tumour stage or patient's reported ancestry). *P* values were FDR corrected to give a FDR *q* value, significance is set to a *q* value of 0.05.

Correlation of clinical features and NKX2-1 amplification. The analysis included 123 consecutive patients with lung adenocarcinoma treated at Brigham and Women's Hospital between January 1997 and December 1999. Fifty-two of these cases had a FISH amplification status that was not assessable (6 cases showed no tumour on the tissue cores and 46 cases had insufficient hybridization). Of the remaining 71 cases, 10 cases had NKX2-1 amplification, 1 had a NKX2-1 deletion, and 60 cases showed no NKX2-1 alteration. All cases for which the NKX2-1 amplification status was not assessable and the one case that showed a NKX2-1 deletion were excluded, bringing the final number of cases included in the analysis to 70.

All cases were histologically confirmed as lung adenocarcinomas. For cases that showed a pure solid growth pattern, mucicarmine and immunohistochemical stains were performed to confirm that the tumour was an adenocarcinoma. Well-differentiated tumours were defined as tumours with a purely bronchioloalveolar growth pattern or mixed tumours with an acinar component with cytologic atypia equivalent to that seen with bronchioloalveolar carcinoma. Poorly differentiated tumours were defined as tumours that showed any amount of solid growth. All other tumours were classified as moderately differentiated. Patient demographics, smoking status, tumour location, type of surgical resection, tumour stage (according to the 6th edition of the American Joint Committee on Cancer system for lung carcinoma) and nodal status were recorded.

Overall survival of patients with NKX2-1 amplification. We excluded from the survival analysis three cases with NKX2-1 amplification and 11 cases that had no NKX2-1 alterations. Exclusion criteria included: cancer was a recurrence; patients received neoadjuvant treatment; patients died within the first 30 days after surgery; and patients had another cancer diagnosed in the 5 years before the diagnosis of lung adenocarcinoma. Survival was plotted by Kaplan–Meier method using the date of resection and date of death or last follow-up.

Sequencing. NKX2-1, MBIP and AUTS2 were sequenced in all 384 TSP lung adenocarcinomas. Primers were designed in an automated fashion using Primer 3 (ref. 36) and characterized by amplification in genomic DNA from three Coriell cell lines. Primers that show an agarose gel band for at least two of the three DNAs were then used for production PCR. Passing primers were arrayed into 384-well PCR plates along with samples and PCR master mix. A total of 5 ng of whole-genome-amplified sample DNA was PCR amplified over 35 cycles in Thermo-Hybrid units, followed by a SAP/Exo clean-up step. NKX2-1 PCR reactions for sequencing contained an addition of 5% DMSO. The resulting purified template is then diluted and transferred to new plates for the sequencing reaction. After cycling (also performed on Thermo-Hybrid), the plates are cleaned up with an ethanol precipitation, re-hydrated and detected on an ABI 3730xl DNA analyser (Applied Biosystems). Output from the detectors is transferred back to the directed sequencing platform's informatics pipeline. SNPs and/or mutations are then identified using three mutation-detecting algorithms in parallel: PolyPhred³⁷ and PolyDHAN (D. Richter *et al.*, manuscript in preparation), which are bundled into the in-house software package SNP Compare, and the commercially available Mutation Surveyor (SoftGenetics, LLC.). Candidates were filtered to remove silent variants, intronic variants (with the exception of potential splice site mutations) and validated SNPs registered in dbSNP or confirmed as SNPs in our previous experiments.

Mutation validation by genotyping. Homogeneous mass extension (hME) genotyping for validation of sequencing candidates was performed in 96-well plates with up to 7-plex reactions. PCR was performed with final concentrations of 0.83 mM dNTPs, 1.56× of 10× buffer, 3.38 mM MgCl₂, 0.03 U μl⁻¹ HotStar Taq (Qiagen), 0.10 μM PCR primers. Thermocycling was performed at 92 °C for 15 min, followed by 45 cycles of 92 °C for 20 s, 56 °C for 30 s and 72 °C for 1 min, with an additional extension at 72 °C for 3 min. Shrimp alkaline phosphatase (SAP) clean-up was performed using a master mix made up of 0.5× buffer and SAP. Reactions were performed at 34 °C for 20 min, 85 °C for 5 min

and then held at 4 °C. After the SAP clean-up, hME reaction was performed using thermosequase and final concentrations of 0.06 mM sequenom termination mix (specific to the pool being used), and 0.64 μM extension primer. Reactions were cycled at 94 °C for 2 min, followed by 55 cycles of 94 °C for 5 s, 52 °C for 5 s and 72 °C for 5 s. Samples were then put through a resin clean-up step, then the purified primer extension reaction was loaded onto a matrix pad (3-hydroxypropylcolonic acid) of a SpectroCHIP (Sequenom) and detected by a Bruker Biflex III MALDI-TOF mass spectrometer (SpectroREADER, Sequenom).

PTPRD mutation discovery and validation. The PTPRD gene was sequenced in 188 lung adenocarcinoma samples. Sequence traces (reads) were aligned to human reference sequence using cross-match. PolyPhred³⁷ and PolyScan were used to predict SNPs and insertions/deletions. Identified SNPs were validated using the Illumina Goldengate assay. ENST00000356435 is the transcript used for annotating the mutations. Both synonymous and non-synonymous candidates were identified, but only non-synonymous mutations were validated.

Tissue microarray FISH (TMA-FISH). A Biotin-14-dCTP-labelled BAC clone RP11-1083E2 (conjugated to produce a red signal) was used for the NKX2-1 probe and a Digoxin-dUTP labelled BAC clone RP11-72J8 (conjugated to produce a green signal) was used for the reference probe. Tissue hybridization, washing and colour detection were performed as described previously^{7,38}. NKX2-1 amplification by FISH was assessed using a total of 935 samples (represented by 2,818 tissue microarray cores).

The BAC clones were obtained from the BACPAC Resource Center, Children's Hospital Oakland Research Institute (CHORI, Oakland, California, USA). Before tissue analysis, the integrity and purity of all probes were verified by hybridization to metaphase spreads of normal peripheral lymphocytes. The samples were analysed under a ×60 oil immersion objective using an Olympus BX-51 fluorescence microscope equipped with appropriate filters, a CCD (charge-coupled device) camera and the CytoVision FISH imaging and capturing software (Applied Imaging). Semi-quantitative evaluation of the tests was independently performed by two evaluators (S.P. and L.A.J.); at least 100 nuclei for each case were analysed when possible. Cases with significant differences between the two independent evaluations were referred by a third person (M.A.R.). The statistical analysis was performed using SPSS 13.0 for Windows (SPSS Inc.) with a significance level of 0.05.

Cell lines and cell culture conditions. NCI-H2009 (ref. 39), NCI-H661 (ref. 40), NCI-H1975 (ref. 39) and HCC1171 (ref. 8) have been previously described. A549 cells were purchased from American Type Culture Collection. NSCLC cells were maintained in RPMI growth media consisting of RPMI 1640 plus 2 mM L-glutamine (Mediatech) supplemented with 10% fetal bovine serum (Gemini Bio-Products), 1 mM sodium pyruvate, and penicillin/streptomycin (Mediatech).

RNAi knockdown. shRNA vectors targeted against NKX2-1, MBIP and GFP were provided by TRC (The RNAi Consortium). The sequences targeted by the NKX2-1 shRNAs are as follows: shNKX2-1a (TRCN0000020449), 5'-CGCTTGTAATACCAGGATTT-3', and shNKX2-1b (TRCN0000020453), 5'-TCCGTCTCAGTGTCTGACAT-3'. The sequences targeted by the MBIP shRNA and GFP shRNA are 5'-CCACCGGAAGGAAGATTATT-3' (TRCN0000003069) and 5'-GCAAGCTGACCCTGAAGTTCAT-3', respectively. Lentiviruses were made by transfection of 293T packaging cells with a three plasmid system^{41,42}. Target cells were incubated with lentiviruses for 4.5 h in the presence of 8 μg ml⁻¹ polybrene. After the incubation, the lentiviruses were removed and cells were fed fresh medium. Two days after infection, puromycin (0.75 μg ml⁻¹ for NCI-H1975, 1.0 μg ml⁻¹ for NCI-H661, 1.5 μg ml⁻¹ for NCI-H2009, 1.0 μg ml⁻¹ for NCI-H661 and 2.0 μg ml⁻¹ for A549 and HCC1171) was added. Cells were grown in the presence of puromycin for 3 days or until all of the non-infected cells died. Twenty-five micrograms of total cell lysates prepared from the puro-selected cell lines was analysed by western blotting using anti-NKX2-1 polyclonal antibody (Santa Cruz Biotechnology), anti-MBIP polyclonal antibody (Proteintech Group, Inc.) and anti-vinculin monoclonal antibody (Sigma).

Soft agar anchorage-independent growth assay. NCI-H2009 (1 × 10⁴), NCI-H661 (2.5 × 10⁴), A549 (3.3 × 10³), NCI-H1975 (5 × 10⁴) or HCC1171 (1 × 10⁴) cells expressing shRNAs targeting NKX2-1, MBIP or GFP were suspended in a top layer of RPMI growth media and 0.4% Noble agar (Invitrogen) and plated on a bottom layer of growth media and 0.5% Noble agar in 35-mm wells. Soft agar colonies were counted 3–4 weeks after plating. The data are derived from two independent experiments unless otherwise noted and are graphed as the percentage of colonies formed relative to the shGFP control cells (set to 100%) ± 1 standard deviation of the triplicate samples. *P* values between shGFP and shNKX2-1 or shMBIP samples were calculated using a *t*-test.

Cell proliferation assays. NCI-H2009 (500 cells per well), A549 (400 cells per well) and NCI-H661 (600 cells per well) cells expressing shRNAs targeting NKX2-1, MBIP or GFP were seeded in 6 wells in a 96-well plate. Cell viability

was determined at 24-h time points for a total of 4 days using the WST-1-based colorimetric assay (Roche Applied Science). The percentage of cell viability is plotted for each cell line ± 1 standard deviation of the reading from six wells, relative to day 0 readings. Experiments were performed two or more times and a representative experiment is shown.

32. Reich, M. *et al.* GenePattern 2.0. *Nature Genet.* **38**, 500–501 (2006).
33. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high-density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
34. Iafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nature Genet.* **36**, 949–951 (2004).
35. Lin, M. *et al.* dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* **20**, 1233–1240 (2004).
36. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386 (2000).
37. Nickerson, D. A., Tobe, V. O. & Taylor, S. L. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res.* **25**, 2745–2751 (1997).
38. Rubin, M. A. *et al.* Overexpression, amplification, and androgen regulation of TPD52 in prostate cancer. *Cancer Res.* **64**, 3814–3822 (2004).
39. Phelps, R. M. *et al.* NCI-Navy Medical Oncology Branch cell line data base. *J. Cell. Biochem., Suppl.* **24**, 32–91 (1996).
40. Banks-Schlegel, S. P., Gazdar, A. F. & Harris, C. C. Intermediate filament and cross-linked envelope expression in human lung tumor cell lines. *Cancer Res.* **45**, 1187–1197 (1985).
41. Naldini, L. *et al.* *In vivo* gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272**, 263–267 (1996).
42. Zufferey, R., Nagy, D., Mandel, R. J., Naldini, L. & Trono, D. Multiply attenuated lentiviral vector achieves efficient gene delivery *in vivo*. *Nature Biotechnol.* **15**, 871–875 (1997).

Identification of protein pheromones that promote aggressive behaviour

Pablo Chamero^{1*}, Tobias F. Marton^{1*}, Darren W. Logan¹, Kelly Flanagan¹, Jason R. Cruz¹, Alan Saghatelian³, Benjamin F. Cravatt² & Lisa Stowers¹

Mice use pheromones, compounds emitted and detected by members of the same species, as cues to regulate social behaviours such as pup suckling, aggression and mating¹. Neurons that detect pheromones are thought to reside in at least two separate organs within the nasal cavity: the vomeronasal organ (VNO) and the main olfactory epithelium (MOE)². Each pheromone ligand is thought to activate a dedicated subset of these sensory neurons. However, the nature of the pheromone cues and the identity of the responding neurons that regulate specific social behaviours are largely unknown. Here we show, by direct activation of sensory neurons and analysis of behaviour, that at least two chemically distinct ligands are sufficient to promote male–male aggression and stimulate VNO neurons. We have purified and analysed one of these classes of ligand and found its specific aggression-promoting activity to be dependent on the presence of the protein component of the major urinary protein (MUP) complex, which is known to comprise specialized lipocalin proteins bound to small organic molecules^{1,3,4}. Using calcium imaging of dissociated vomeronasal neurons (VNs), we have determined that the MUP protein activates a sensory neuron subfamily characterized by the expression of the G-protein G_{α_o} subunit (also known as G_{nao}) and $Vmn2r$ putative pheromone receptors (V2Rs). Genomic analysis indicates species-specific co-expansions of MUPs and V2Rs, as would be expected among pheromone-signalling components. Finally, we show that the aggressive behaviour induced by the MUPs occurs exclusively through VNO neuronal circuits. Our results substantiate the idea of MUP proteins as pheromone ligands that mediate male–male aggression through the accessory olfactory neural pathway.

Male–male territorial aggression in mice is a robust, innate, social behaviour. However, the aggression-promoting pheromone(s) and the responding neural circuits that mediate aggression are unknown. Castrated males no longer produce the aggression pheromone and fail to stimulate aggressive behaviour from recipient males. However, whole urine from intact males is sufficient to promote aggression when swabbed on the backs of castrated animals⁵, providing a bioassay for the identification of urinary pheromones (Fig. 1a). We used this behavioural assay to determine which components of urine act as pheromones that cause aggression. We first fractionated male urine over size-separation columns and tested these fractions in the castrated-male bioassay. We found that fractions comprising molecules of low (LMW; less than 3 kilodaltons (kDa)) and high (HMW; greater than 10 kDa) molecular mass both contained aggression-promoting activity (Fig. 1b). The behavioural characteristics of the observed aggression promoted by LMW and HMW fractions were indistinguishable from each other and from the behaviour promoted

by whole urine (data not shown). These findings suggest that at least two distinct molecules promote aggression.

To assay pheromone activity further, we established an *ex vivo* system using primary sensory neurons suitable for screening many heterogeneous cells for biologically active ligands. Our previous studies revealed that VNs are required for the aggression response, because mice lacking the primary sensory transduction channel, TRPC2, are unable to detect and respond to the aggression-promoting pheromone (Fig. 1a)^{6–8}. We found that dissociated primary VNs loaded with fura-2 responded to male whole urine with robust and reproducible intracellular Ca^{2+} transients (Fig. 1c). A battery of controls and a dose–response curve (Figs 1f and 2e, and Supplementary Fig. 1), based on the molecular characteristics and physiology of VNs, show the response specificity of urine and, importantly, establish that dissociated VNs provide a biological platform to analyse the activity of potential pheromone ligands.

To investigate the LMW and HMW fractions further, we analysed the activation of dissociated VNs by each size fraction. VNs are a highly heterogeneous population, with each neuron expressing one of approximately 250 different G-protein coupled receptors (GPCRs)⁹, providing a mechanism for individual neurons to respond to different compounds. We next determined whether the LMW and HMW fractions activated distinct or overlapping populations of dissociated VNs. Our calcium imaging method allows us to record calcium transients in response to repetitive exposure of multiple substances at a resolution of the single cell; this enables us to determine precisely which ligands are biologically active as well as the response profile of individual neurons. When assayed, one population of the responding cells was activated by the HMW stimulus whereas a second distinct population showed calcium transients in response to the LMW ligands (Fig. 1d–f). This indicates that two chemically distinct ligands activate separate subsets of VNs. When considered with the bioassay, it suggests that at least two populations of neurons are capable of detecting urinary aggression pheromones and that each are sufficient to promote male–male aggression.

There are very few HMW components in mouse urine^{10,11}; none has been identified as a pheromone. Therefore we chose to focus our subsequent purification and characterization on only the robust HMW bioactivity. We used anion-exchange fast protein liquid chromatography (FPLC) to separate the HMW components into 40 fractions over a 0–1 M NaCl gradient (Supplementary Fig. 2). Only five fractions (fractions 15–19) induced calcium transients in VNs. This activity overlapped with and accounted for all the HMW activity. Polyacrylamide gel electrophoresis (PAGE) revealed that the five active fractions contained proteins of 19–24 kDa, which can be further resolved into four major bands by isoelectric focusing (Fig. 2a);

¹Department of Cell Biology, ²Department of Chemistry, The Scripps Research Institute, La Jolla, California 92030, USA. ³Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

*These authors contributed equally to this work.

these features closely mirror the size and isoelectric point (pI) of MUPs⁴. Indeed, western blot (Fig. 2a) and electrospray ionization mass spectrometry (ESIMS) of these five fractions confirmed their identities as MUPs. Of the 20 identified MUP-encoding genes arrayed in the genome, males are known to express unique combinations of four to six MUPs in a strain-dependent pattern¹². Figure 2a identifies the four MUPs excreted by C57BL/6J males. Importantly, we did not detect any other proteins in these purified fractions. When used in the behavioural assay, we observed that the MUPs purified by FPLC are sufficient to mediate robust male–male territorial aggression (Fig. 2f).

MUPs are β -barrel in structure, bearing a central hydrophobic binding pocket that has been shown to carry small organic ligands¹³. Gas chromatography followed by mass spectrometry (GC-MS) revealed that our isolated MUPs primarily bind 2-*s*-butyl-4,5-dihydrothiazole

(SBT) (Supplementary Fig. 3). Previous studies have implicated SBT as a pheromone¹⁴ capable of activating a subset of VNs¹⁵; however, the role of MUP itself without ligands, MUP protein, remains elusive^{4,16,17}.

To investigate the function of the MUP protein further, we first eliminated the protein component of the purified MUP complex by protease digestion. This treatment abolished the ability of the purified complex both to activate VNs and to promote aggressive behaviour (Fig. 2b, f). Next, we investigated the specificity of the small-molecule ligand in promoting aggression. We tested synthetic SBT in our bioassay and found no aggression-promoting activity (Fig. 2f). The binding affinities and infinite characteristics of potential small-molecule ligands preclude the definitive dissociation of all ligands from MUPs. Therefore, we incubated fractions 15–19 with

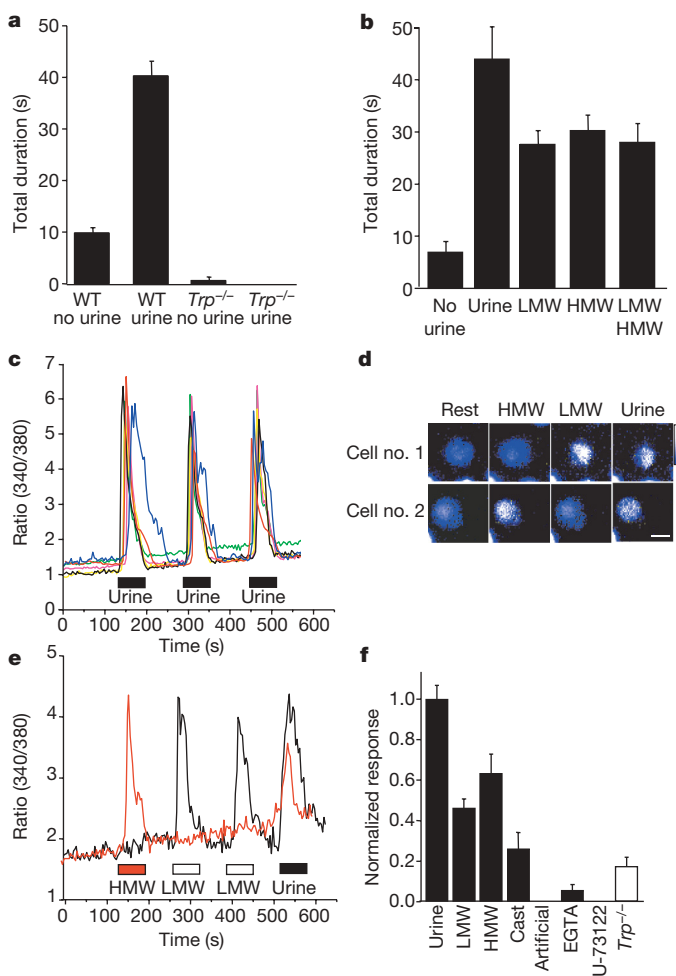


Figure 1 | Male urine contains two aggression pheromones. **a**, Male urine swabbed on castrated mice stimulates aggression ($P < 0.0001$; Student's *t*-test) in WT (73 trials/36 animals) but not *Trpc2*^{-/-} mutants (36 trials/6 animals, mean \pm s.e.m.). Aggression, total attack duration. **b**, Aggression with urine, LMW, HMW or both ($N = 16$ trials/6 animals each; urine/HMW $P = 0.1$, urine or HMW/no urine $P < 0.0001$). Error bars, s.e.m. **c**, Repetitive application of male urine (1:300) induced Ca^{2+} transients in dissociated VNs. Six representative traces. **d**, Fura-2 ratio images of two VNs of the same experiment. Cell no. 1 responds to LMW, cell no. 2 to HMW. Scale bar, 10 μm . Pseudocolour: low (blue) to high (white) $[\text{Ca}^{2+}]$. **e**, Separate populations of VNs are activated by LMW (black) and HMW (red) fractions. **f**, Summary of VN activation (mean \pm s.e.m. normalized to the urine response): WT (black bars) stimulated with urine, 1,951 of 28,289 cells; LMW, 548 of 17,260 cells; HMW, 885 of 21,096 cells; common to both, 80 of 12,679; 'castrated' urine, 44 of 2,153; artificial urine²³, 0 of 1,224; EGTA, calcium-free media, 9 of 2,426; PLC inhibitor U-73122 (50 μM), 0 of 2,205. *Trp*^{-/-} (white bar), *Trpc2*^{-/-} VNs in response to urine, 38 of 3,312 (ref. 24).

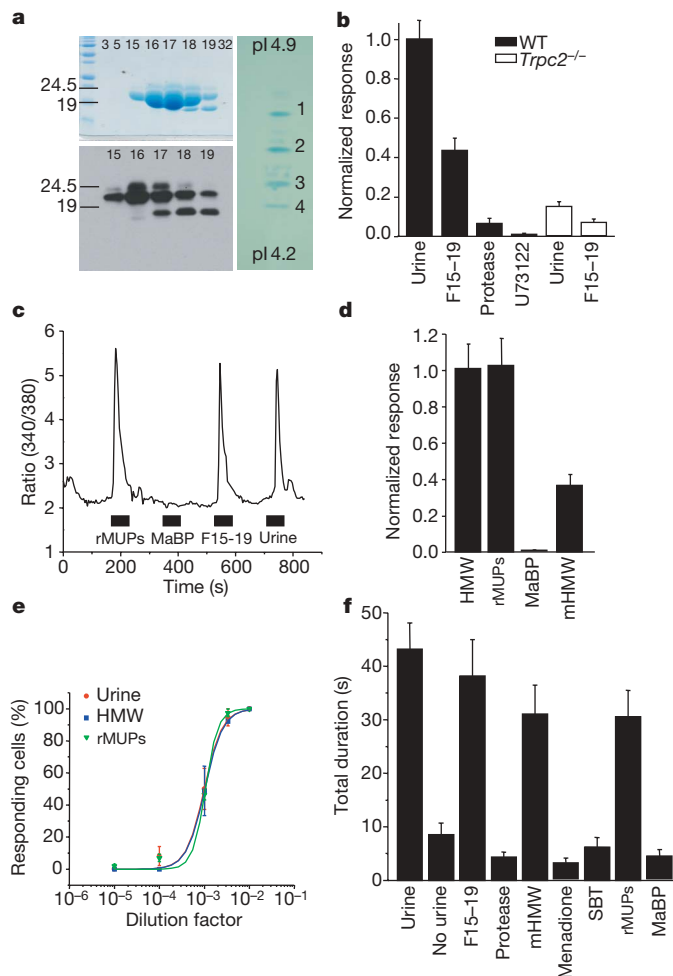


Figure 2 | HMW aggression activity is dependent on MUPs. **a**, Purification of bioactivity. Top, PAGE of FPLC fractions (Supplementary Fig. 2). Bottom, anti-MUP western blot. Right, isoelectric focusing gel (pI range 4.1–4.9) of the fractions (F15–19); accession numbers: 1, AAI00587; 2, CAM19799; 3, AAH13649; 4, AAH19965. **b**, Calcium imaging of VNs. F15–19 activated one half (72 of 1,220 cells) of the same VNs activated by whole urine (133 of 1,220 cells). Proteinase treatment of F15–19 (8 of 1,220 cells) and U-73122 (50 μM ; 2 of 1,563 cells) ablated activity. *Trpc2*^{-/-} VNs show decreased activity; whole urine (28 of 1,737 cells), F15–19 (13 of 1,737 cells)²⁴. **c**, Calcium transients in a single VN induced by rMUPs, F15–19, and whole urine but not maltose-binding protein (MaBP) alone. **d**, VN activation normalized to the HMW response. rMUPs (573 of 4,613 cells); MaBP (1 of 4,613 cells); menadione-displaced HMW (mHMW) (190 of 3,997 cells) and HMW (808 of 6,573 cells). **e**, Dose–response fitted to a sigmoid curve using the Hill equation of VN activation by urine. **f**, Aggressive behaviour measured as total attack duration in resident–intruder assay ($n = 21$ –50). rMUPs/no urine $P = 0.00002$; F15–19/ rMUPs $P = 0.3599$. Error bars in **b**, **d** and **f**, s.e.m.

menadione, to competitively displace MUP small-molecule ligands¹⁸, as analysed by GC–MS (Supplementary Fig. 3). This displaced fraction retains 40% of its original activity, as determined by calcium imaging (Fig. 2d); however, importantly, it retains all of the behavioural aggression-promoting activity (Fig. 2f). This indicates that the MUP protein determines neuronal activation that encodes male–male aggressive behaviour, irrespective of the specificity of its small molecule. Lastly, we prepared the four MUPs excreted in urine from C57BL/6J mice as recombinant maltose–MUP fusion proteins in *Escherichia coli* (rMUPs), and determined by GC–MS that they are not bound with mouse urinary small molecules (data not shown). These pooled rMUPs both induce intracellular calcium transients in VNs and promote aggressive behaviour, demonstrating the functional necessity and sufficiency of the MUP protein as the HMW activity (Figs 2c, d, f). Finally HMW, rMUPs and urine all show similar dose–response activation profiles as analysed by the number of responding cells (Fig. 2e). Together, these data reveal a role for the MUPs without ligands as independent pheromones.

The mouse VNO is composed of two molecularly distinct populations of sensory neurons, defined by the expression of $G\alpha_{i2}$ and $G\alpha_o$, that project to two physically separate domains of the accessory olfactory bulb². The functional significance of these two neuronal classes has yet to be determined. However, the small-molecule ligands alone, such as SBT, are thought to activate the $G\alpha_{i2}$ -expressing neurons¹⁵. Therefore, we aimed to establish the extent to which the MUPs initiate aggression through the activation of one or both classes of VNs. We used calcium imaging followed by immunostaining for $G\alpha_{i2}$ and $G\alpha_o$ to identify the molecular characteristics of those cells activated by MUPs. Figure 3a, b reveals that MUPs specifically activate the $G\alpha_o$ -positive VNs that also express V2R receptors.

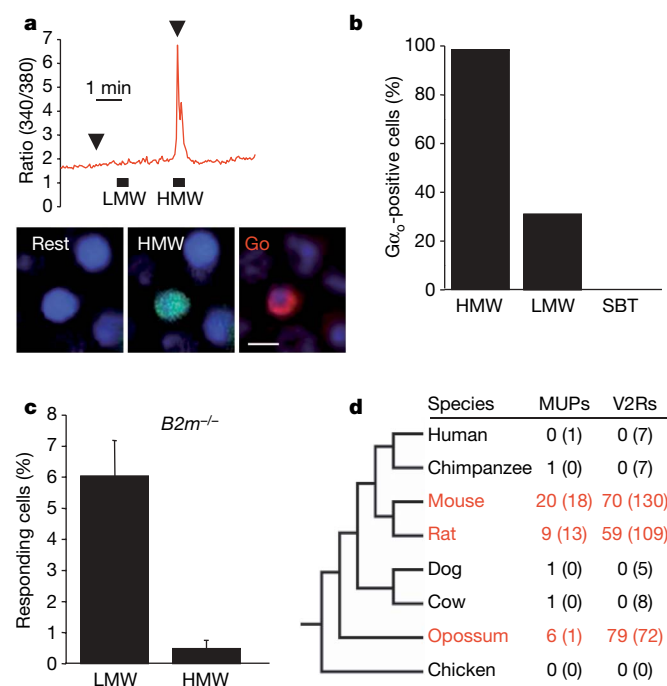


Figure 3 | MUPs activate a subset of VNs that express $G\alpha_o$. **a**, HMW-responsive cell (red trace) is labelled by anti- $G\alpha_o$ immunostaining (red) and DAPI (blue) immediately after Ca^{2+} imaging (right). Fura-2 ratio images (pseudocolour) during rest (left) or HMW activation (middle). Arrows show exact image time. **b**, Percentage of activated cells positive for $G\alpha_o$: HMW 98.6% (75 of 76 cells), LMW 29% (27 of 92), SBT (0 of 4). **c**, $B2m^{-/-}$ VNs activated by LMW (99 of 1,673) and HMW (8 cells to both HMW and LMW). No cells responded only to HMW (mean \pm s.e.m. of six experiments). **d**, Co-expansion of MUP and V2R gene families. The numbers of genes and pseudogenes (brackets) are indicated. V2R data are as reported⁹, with the addition of chimpanzee.

None of the MUP-activated cells was immunoreactive for $G\alpha_{i2}$. Previous studies have shown that $\beta 2$ -microglobulin^{-/-} mice do not properly traffic putative pheromone V2R receptors in the $G\alpha_o$ -expressing neurons and fail to display male–male aggression¹⁹. To assay functionally if the MUPs are indeed signalling through $G\alpha_o$ -expressing neurons, we examined the ability of the purified MUPs to evoke activity in VNs dissociated from the $\beta 2$ -microglobulin-negative ($B2m^{-/-}$) animals. Unlike wild-type (WT) neurons, we found MUP-mediated activity to be abolished in the mutant neurons (Fig. 3c). Together, these results demonstrate that the VNs activated by the MUPs belong to the $G\alpha_o$ subset of VNs. Moreover, the neurons activated by the MUP protein are different to the $G\alpha_{i2}$ -expressing neurons shown to be activated by the MUP-associated small-molecule ligands alone (Fig. 3b)¹⁵. Together, our results indicate that MUPs act as male–male aggression pheromones that specifically stimulate the $G\alpha_o$ -expressing subpopulation of VNs.

Recent comparative genomic and morphological analyses have shown that not all terrestrial vertebrates express markers and functional receptors of the $G\alpha_o$ neurons, including the family of V2R putative pheromone receptors^{9,20}. We analysed sequenced genomes and identified the presence of V2R and MUP gene expansion only in the genomes of rat and mouse, and a parallel expansion of V2R and MUP-like genes in the evolutionary divergent opossum lineage (Fig. 3d and Supplementary Fig. 5). All other mammals studied contain a single, intact MUP gene within the syntenic region, except humans, which have a single pseudogene. The species-specific co-expansion of MUPs and V2Rs further underscores the likelihood that they are both components functioning in species-specific processes, as would be predicted in pheromone communication.

Previous behavioural experiments have found that, like *Trpc2*^{-/-} animals, male mice defective in MOE signalling do not initiate male–male aggression^{21,22}. This prompted us to ask whether the MUP complex was additionally activating MOE neurons. Calcium induced by whole urine and the LMW fraction increases in dissociated MOE neurons; however, we were unable to detect any activation by the purified MUPs (Fig. 4a, b). Our results suggest that the MUP protein mediates male–male aggression exclusively through VNO circuitry. The previously identified necessity of MOE signalling may compose a second, independent pheromone-responsive circuit. Purification and analysis of the LMW aggression-promoting pheromone will enable us to address the nature of this dual processing further.

Behavioural analysis followed by direct VN activation has allowed us to begin to unravel the nature of the aggression-promoting pheromone code. We found that at least two pheromone cues independently promote aggressive behaviour (Fig. 4c). The underlying neuronal logic that promotes an animal's behaviour is not well understood. Several characteristics such as gender, age, status or individuality may be transmitted in the pheromone profile, each acting as equal cues triggering male–male aggression. The MUPs

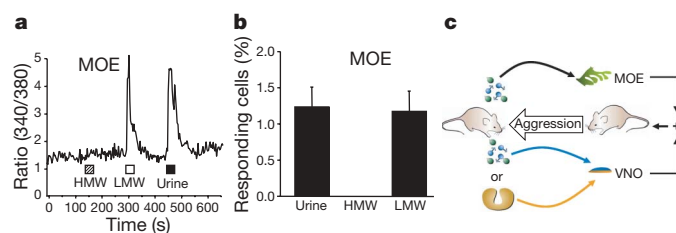


Figure 4 | MUP activation is specific to VNs. **a**, MOE-dissociated cells are not activated by HMW. **b**, Summary of MOE cell activation by urine 1.23% (40 cells), HMW (0 cells), LMW 1.17% (38 cells) of 3,250 total cells sampled (mean \pm s.e.m. of six experiments). **c**, Male–male aggression is mediated by at least two sufficient pheromones: MUPs through $G\alpha_o$ /V2R VNs (orange arrow) and unidentified LMW pheromones that stimulate either the VNO or both the VNO and the MOE (blue and black arrows). Previous genetic experiments indicate that both a functional VNO and MOE are necessary for aggressive behaviour^{6,8,21,22}.

and the unidentified LMW ligands may encode any of these characteristics, independently promoting aggression when encountered by another adult male. Identification of the entire repertoire of aggression-promoting neurons will allow investigation to determine the logic and integration of multiple aggression-promoting circuits that underlie the regulation of behaviour.

METHODS SUMMARY

Calcium imaging. VNs were prepared from male C57Bl/6J mice by dissection followed by dissociation with papain and plating on coverslips coated with concanavalin A. Dissociated VNs were perfused with stimuli, and intracellular calcium was monitored using fura-2/AM (Molecular Probes) in a Zeiss Axiovert 200M inverted microscope. Urine was collected from 8- to 12-week-old C57Bl/6J males and used or further fractionated for behavioural and physiological experiments.

MUP purification. Size fractionation of urine was performed using Centricon filtrating columns (3 kDa and 10 kDa, Millipore). Purification of MUPs from the HMW fraction was completed by using a HiTrap Q HP anion exchange column (GE) fixed to an AKTA FPLC apparatus (Amersham Pharmacia). Isoelectric focusing of MUPs was performed on an LKB 2117 Multiphor II Flatbed Electrophoresis Unit using Immobiline dry-plate gel, pH range 4.2–4.9, and cooled to 10 °C. Protease treatment of FPLC purified MUPs was performed by overnight incubation at 37 °C with proteinase K and papain. Recombinant MUP proteins were generated using the pMAL Protein Fusion and Purification system (New England Biolabs), and normalized to 1× urine by molarity for all calcium-imaging and behavioural experiments.

Genomics. MUP genes were searched in the genome assemblies of the mouse (*Mus musculus*, NCBI m36), rat (*Rattus norvegicus*, RGSC 3.4), human (*Homo sapiens*, NCBI 36), chimpanzee (*Pan troglodytes*, PanTro 2.1), dog (*Canis familiaris*, Canfam 2.0), cow (*Bos taurus*, Btau 3.1), opossum (*Monodelphis domestica*, Mondom 4.0) and chicken (*Gallus gallus*, WASHUC2) using a modification of the methods used by Shi and Zhang (2007)⁹.

Behaviour. The resident–intruder assay was performed as previously described using 40 µl of stimulus normalized to 1× urine with protocols approved by the Institutional Animal Care and Use Committee⁶.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 July; accepted 12 October 2007.

1. Stowers, L. & Marton, T. What is a pheromone? Mammalian pheromones reconsidered. *Neuron* **46**, 699–702 (2005).
2. Dulac, C. & Torello, A. T. Molecular detection of pheromone signals in mammals: from genes to behaviour. *Nature Rev. Neurosci.* **4**, 551–562 (2003).
3. Cavaggioni, A. & Mucignat-Caretta, C. Major urinary proteins, α_{2U} -globulins and aphrodisin. *Biochim. Biophys. Acta* **1482**, 218–228 (2000).
4. Flower, D. R. The lipocalin protein family: structure and function. *Biochem. J.* **318**, 1–14 (1996).
5. Mugford, R. A. & Nowell, N. W. Pheromones and their effect on aggression in mice. *Nature* **226**, 967–968 (1970).
6. Stowers, L. *et al.* Loss of sex discrimination and male–male aggression in mice deficient for TRP2. *Science* **295**, 1493–1500 (2002).
7. Bean, N. J. Modulation of agonistic behavior by the dual olfactory system in male mice. *Physiol. Behav.* **29**, 433–437 (1982).

8. Leypold, B. G. *et al.* Altered sexual and social behaviors in trp2 mutant mice. *Proc. Natl Acad. Sci. USA* **99**, 6376–6381 (2002).
9. Shi, P. & Zhang, J. Comparative genomic analysis identifies an evolutionary shift of vomeronasal receptor gene repertoires in the vertebrate transition from water to land. *Genome Res.* **17**, 166–174 (2007).
10. Schwende, F. J. *et al.* Urinary volatile constituents of the house mouse, *Mus musculus*, and their endocrine dependency. *J. Chem. Ecol.* **12**, 277–296 (1986).
11. Hastie, N. D., Held, W. A. & Toole, J. J. Multiple genes coding for the androgen-regulated major urinary proteins of the mouse. *Cell* **17**, 449–457 (1979).
12. Robertson, D. H. *et al.* Molecular heterogeneity of urinary proteins in wild house mouse populations. *Rapid Commun. Mass Spectrom.* **11**, 786–790 (1997).
13. Timm, D. E. *et al.* Structural basis of pheromone binding to mouse major urinary protein (MUP-I). *Protein Sci.* **10**, 997–1004 (2001).
14. Novotny, M., Harvey, S., Jemiolo, B. & Alberts, J. Synthetic pheromones that promote inter-male aggression in mice. *Proc. Natl Acad. Sci. USA* **82**, 2059–2061 (1985).
15. Leinders-Zufall, T. *et al.* Ultrasensitive pheromone detection by mammalian vomeronasal neurons. *Nature* **405**, 792–796 (2000).
16. Hurst, J. L., Robertson, D. H. L., Tolladay, U. & Beynon, R. J. Proteins in urine scent marks of male house mice extend the longevity of olfactory signals. *Anim. Behav.* **55**, 1289–1297 (1998).
17. Hurst, J. L. *et al.* Individual recognition in mice mediated by major urinary proteins. *Nature* **414**, 631–634 (2001).
18. Xia, J. *et al.* Urinary pheromones promote ERK/Akt phosphorylation, regeneration and survival of vomeronasal (V2R) neurons. *Eur. J. Neurosci.* **24**, 3333–3342 (2006).
19. Loconto, J. *et al.* Functional expression of murine V2R pheromone receptors involves selective association with the M10 and M1 families of MHC class Ib molecules. *Cell* **112**, 607–618 (2003).
20. Takigami, S., Mori, Y., Tanioka, Y. & Ichikawa, M. Morphological evidence for two types of mammalian vomeronasal system. *Chem. Senses* **29**, 301–310 (2004).
21. Mandiyan, V. S., Coats, J. K. & Shah, N. M. Deficits in sexual and aggressive behaviors in *Cnga2* mutant mice. *Nature Neurosci.* **8**, 1660–1662 (2005).
22. Wang, Z. *et al.* Pheromone detection in male mice depends on signaling through the type 3 adenylyl cyclase in the main olfactory epithelium. *J. Neurosci.* **26**, 7375–7379 (2006).
23. Holy, T. E., Dulac, C. & Meister, M. Responses of vomeronasal neurons to natural stimuli. *Science* **289**, 1569–1572 (2000).
24. Lucas, P., Ukhanov, K., Leinders-Zufall, T. & Zufall, F. A diacylglycerol-gated cation channel in vomeronasal neuron dendrites is impaired in TRPC2 mutant mice: mechanism of pheromone transduction. *Neuron* **40**, 551–561 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank F. Papes, U. Mueller, A. Patapoutian, K. Baldwin and C. Zucker for discussions and critical reading of the manuscript. This work was supported by funding from the NIDCD, Pew Charitable Trust, Skaggs Institute, and Helen Dorris Foundation (to L.S.), a National Institute on Deafness and Other Communication Disorders (NIDCD) pre-doctoral fellowship (T.F.M.) and The Basque Government Post-Doctoral Research Fellowship (P.C.).

Author Contributions The behavioural analysis was performed by P.C., J.R.C. and L.S.; calcium imaging was done by P.C. and T.F.M.; and biochemical purification of MUPs by T.F.M. and advised by B.F.C. SBT synthesis was performed by A.S. and B.F.C.; construct preparation and immunostaining by P.C., T.F.M. and K.F.; and genomic analysis by D.W.L. All authors participated in data analysis and writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to L.S. (stowers@scripps.edu).

METHODS

Cell preparation. Male 8- to 12-week-old C57BL/6J mice were used for all the experiments. The VNO was removed to dissect the epithelium. The tissue was incubated for 20 min at 37 °C in cation-free 0.22 units ml⁻¹ papain, 5.5 mM cysteine-HCl and 10 U ml⁻¹ DNase I in PBS. The papain was inactivated with 10% FBS containing D-MEM and the dissociated cells were plated on 12 mm round coverslips coated with concanavalin-A. For dissociated MOE cells, the whole MOE was first dissected and dissociated in 1 ml PBS containing 40 mM urea, 0.22 U ml⁻¹ papain and 10 U ml⁻¹ DNase I for 20 min at 37 °C. β 2-microglobulin^{-/-} mice were purchased from Taconic.

Calcium imaging. Intracellular Ca²⁺ was monitored using fura-2/AM (5 μ M, Molecular Probes) in a Zeiss Axiovert 200M inverted microscope with a 20 \times fluar 0.75 objective lens. Cells were loaded with HBSS supplemented with 10 mM HEPES, and incubated for 30–60 min at room temperature. Coverslips were placed in a temperature-controlled (37 °C) laminar-flow perfusion chamber (Warner Instrument Corp.) and constantly perfused with HEPES-buffered HBSS. Fura-2-loaded cells were excited with wavelengths alternating between 340 and 380 nm, and light of wavelength greater than 540 nm was captured with an Orca-ER camera (Hamamatsu). After subtraction of background fluorescence, the ratio of fluorescence intensity at the two wavelengths was calculated and analysed using MetaFluor (Universal Imaging Corporation) and NIH Image J. Urine was diluted 1:300 in HBSS; test fractions and purified MUPs were normalized to 1 \times urine and then diluted 1:300 before experimentation.

Urine fractionation. C57BL/6J male mice of 8–12 weeks age were used as the source of urine. Between 0.5 and 1 ml of urine was size fractionated by centrifugation (5,000g, 30 min), using Centricon molecular weight cut-off filtrating columns (3 kDa and 10 kDa, Millipore). The first centrifugation flowthrough was collected as the LMW fraction. The HMW retentate was washed with one volume of PBS three times and re-concentrated to reach the same initial concentration of urine. The composition of artificial urine was (in mM): 120 NaCl, 40 KCl, 20 NaH₄OH, 4 CaCl₂, 2.5 MgCl₂, 15 NaH₂PO₄, 20 NaHSO₄, 333 Urea, pH 7.4 (ref. 23).

Protease treatment of MUPs. The four pooled FPLC-purified MUPs were incubated overnight at 37 °C with a protease cocktail of 0.22 U ml⁻¹ proteinase K and 0.22 U ml⁻¹ papain. PAGE showed the digestion to be 95% complete. The digested proteins were spun in a Centricon 3 kDa molecular weight cut-off filtration column to remove undigested and partly digested MUPs.

IEF. Isoelectric focusing of C57BL/6J MUPs was performed on a LKB 2117 Multiphor II Flatbed Electrophoresis Unit using an Immobiline dry-plate gel, pH range 4.2–4.9, and cooled to 10 °C. Male C57BL/6J urine was de-salted over a G-50 Microspin Column (GE) and 5 μ l of sample was applied directly to the gel. Samples were loaded into the gel at 200 V, 5 mA and 15 W for 200 V h. The gel was electrophoresed at 3,500 V, 5 mA and 15 W for 14.8 kV h and then fixed and stained with Coomassie brilliant blue.

Behaviour. C57BL/6J male mice (8–12 weeks old) were isolated for one week. The mice were exposed to castrated adult mice swabbed with 40 μ l of test solution (1 \times male urine; fractions and FPLC-purified MUPs were normalized to 1 \times urine) and assayed for 10 min. Tests took place in the home cages of isolated mice, and at least 48 h was allowed before a new test was conducted. Tests were videotaped and analysed at quarter speed using Observer software (Noldus Technology) to measure aggression parameters including tail rattling, biting, chasing, cornering, tumbling and kicking. Total duration was defined as the total duration of aggressive contact behaviour consisting of kicking, biting, wrestling or tumbling. One round of urine and no-urine controls was performed with each resident mouse before and after sample testing.

Production of recombinant MUP. Recombinant MUP protein was produced using the pMAL Protein Fusion and Purification System (New England Biolabs). Full-length MUP complementary DNAs (cDNAs) corresponding to the four C57BL/6J MUPs expressed in urine were cloned from a male C57BL/6J liver cDNA library and subcloned into pMAL bacterial expression vector pMAL-c2X. The starter culture was diluted into 1 litre LB/AMP/2% glucose, grown for 1 h at 37 °C followed by 2 h of induction with 0.3 mM isopropyl β -D-1-thiogalactopyranoside (IPTG). Cells were centrifuged at 4,000g, 20 min and resuspended in 25 ml Column buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA) plus protease inhibitors (Roche) and incubated for 30 min on ice with 1 mg ml⁻¹ lysozyme. The sample was sonicated and then centrifuged (9,000g for 30 min). The supernatant was incubated overnight at 4 °C with 2 ml bed volume amylose resin and subsequently washed three times with 50 ml cold column buffer. rMUPs were eluted with 2 ml column buffer plus 25 mM maltose for 2 h at room temperature. rMUPs were assayed by SDS-PAGE. All rMUPs were pooled and normalized to 1 \times urine for behavioural analysis and further diluted 1:300 for calcium imaging.

Dose-response curve. For all calcium imaging experiments, stimuli were normalized to the concentration of MUPs in 1 \times urine (20 mg ml⁻¹ as determined by Bradford assay) and then diluted 1:300 in Hanks/HEPES buffer. The four rMUP fusion proteins were pooled together using the estimation that each MUP is present in urine at one quarter of the concentration (5 mg ml⁻¹) of all MUPs. The pooled rMUPs were normalized to 1 \times urine by molarity. The dose-response curve was generated by presenting the stimuli (urine, HMW or pooled rMUPs) to VNs serially in the following dilutions: 1:100,000, 1:10,000, 1:1,000, 1:300, 1:100. The number of responding cells was counted for each dilution and normalized to the maximum number of responding cells observed. The dose-response was fitted to a sigmoid curve by using the Hill equation. Urine effector concentration for half-maximum response (EC₅₀) = 0.00099, slope = 2.15, n = 135 cells in four experiments; HMW EC₅₀ = 0.001, slope = 2.14, n = 52 cells in two experiments; rMUPs EC₅₀ = 0.0011, slope = 3.0, n = 209 cells in four experiments.

Adaptive immunity maintains occult cancer in an equilibrium state

Catherine M. Koebel¹, William Vermi^{1,2}, Jeremy B. Swann^{3,4}, Nadeen Zerafa³, Scott J. Rodig⁵, Lloyd J. Old⁶, Mark J. Smyth^{3,4*} & Robert D. Schreiber^{1*}

The capacity of immunity to control and shape cancer, that is, cancer immunoediting, is the result of three processes^{1–8} that function either independently or in sequence⁹: elimination (cancer immunosurveillance, in which immunity functions as an extrinsic tumour suppressor in naive hosts); equilibrium (expansion of transformed cells is held in check by immunity); and escape (tumour cell variants with dampened immunogenicity or the capacity to attenuate immune responses grow into clinically apparent cancers). Extensive experimental support now exists for the elimination and escape processes because immunodeficient mice develop more carcinogen-induced and spontaneous cancers than wild-type mice, and tumour cells from immunodeficient mice are more immunogenic than those from immunocompetent mice. In contrast, the equilibrium process was inferred largely from clinical observations, including reports of transplantation of undetected (occult) cancer from organ donor into immunosuppressed recipients¹⁰. Herein we use a mouse model of primary chemical carcinogenesis and demonstrate that equilibrium occurs, is mechanistically distinguishable from elimination and escape, and that neoplastic cells in equilibrium are transformed but proliferate poorly *in vivo*. We also show that tumour cells in equilibrium are unedited but become edited when they spontaneously escape immune control and grow into clinically apparent tumours. These results reveal that, in addition to destroying tumour cells and sculpting tumour immunogenicity, the immune system of a naive mouse can also restrain cancer growth for extended time periods.

To assess whether an equilibrium state can develop during primary tumorigenesis, groups of sex- and age-matched wild-type C57BL/6 or 129/SvEv mice were injected with a single low dose of the chemical carcinogen 3'-methylcholanthrene (MCA) (Fig. 1a) and monitored for the appearance of progressively growing sarcomas for the next 200 or 230 days, after which new tumour formation decreases precipitously in C57BL/6 and 129/SvEv mice, respectively (Supplementary Fig. 1). Mice developing progressively growing sarcomas during the aforementioned times were removed from the experiments. The remaining mice, which often displayed small stable masses at the site of MCA injection, were placed on weekly injections of either control monoclonal antibody or monoclonal antibodies that deplete or block specific immunological components, and thereafter monitored for the appearance of progressively growing tumours.

When two cohorts of wild-type C57BL/6 mice were treated with 25 µg MCA, 3/16 (19% cohort 1, Fig. 1b) and 4/19 (21%, cohort 2, Fig. 1c) developed progressively growing sarcomas by day 200 and were removed from the experiment. When the remaining mice in

cohort 1 were treated with control immunoglobulin, none developed additional tumours (Fig. 1b). In contrast, when the remaining mice in cohort 2 were treated with a mixture of monoclonal antibodies that deplete CD4⁺ and CD8⁺ cells and neutralize interferon-gamma (IFNγ), 60% (9/15) developed progressively growing sarcomas at the MCA injection site (Fig. 1c).

Variations of this experiment were repeated several times in two laboratories, using mice with either homogeneous C57BL/6 or 129/SvEv backgrounds. The results were in complete agreement and revealed that anti-CD4/CD8, anti-IFNγ, or anti-IL-12p40 (which is critical for IFNγ production) were equally effective in inducing sarcoma outgrowth and were not more effective when used in combination (Fig. 1d, and Supplementary Fig. 2). In total, of 187 mice treated with low-dose MCA, 86 (46%) developed progressively growing sarcomas following depletion of CD4/CD8 cells, IFNγ and/or IL-12—components that participate in adaptive immunity. In contrast, tumour outgrowth did not occur when monoclonal antibodies were used that deplete natural killer cells (anti-NK1.1), block natural killer cell recognition (anti-NKG2D) or inhibit one natural killer cell effector function (anti-TNF-related apoptosis inducing ligand (TRAIL)) (*n* = 55) (Fig. 1d). These data suggest that adaptive immunity plays a particularly important part in preventing late MCA-induced sarcoma outgrowth.

Two possibilities could explain the late tumour outgrowth we observed on ablating adaptive immunity: either the treatment permitted delayed *de novo* transformation or it facilitated expansion of pre-formed occult cancer cells. To explore the former and further substantiate the selective importance of adaptive immunity in preventing late tumour outgrowth, experiments were conducted using mice lacking either recombination-activating gene (*Rag*)-1 or *Rag*2, which possess an intact innate immune compartment but lack lymphocytes that mediate adaptive immunity (that is, T cells, natural killer T cells and B cells). In stark contrast to the results obtained with wild-type mice, very few late-forming tumours were observed in *Rag1*^{−/−} C57BL/6 mice exposed to 5 µg MCA (a dose that induces tumour outgrowth comparable to that in wild-type mice exposed to 25 µg MCA) and subsequently treated on day 200 with control immunoglobulin, anti-CD4/-CD8/-IFNγ (Fig. 2a, b), or anti-NK1.1 (Fig. 2c). Very few late-forming tumors were also observed in *Rag2*^{−/−} 129/SvEv mice treated with 25 µg MCA (so as to provide a carcinogenic insult comparable to that used for wild-type mice) and subsequently treated on day 230 with control immunoglobulin or anti-IFNγ (to assess the effects of natural-killer-cell-produced IFNγ on cells of innate immunity) (Fig. 2d). These results show that sarcoma formation in immunodeficient *Rag*^{−/−} mice was essentially

¹Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63110, USA. ²Department of Pathology, University of Brescia/Spedali Civili di Brescia, Brescia 25123, Italy. ³Cancer Immunology Program, Sir Donald and Lady Trescowthick Laboratories, Peter MacCallum Cancer Centre, East Melbourne, Victoria 3002, Australia. ⁴Department of Pathology, University of Melbourne, Parkville, Victoria 3010, Australia. ⁵Department of Pathology, Brigham and Women's Hospital Harvard Medical School, Boston, Massachusetts 02115, USA. ⁶Ludwig Institute for Cancer Research at Memorial Sloan Kettering Cancer Center, New York 10021, USA.

*These authors contributed equally to this work.

complete within 200 days of MCA exposure. In addition, the mean time to tumour formation in MCA-treated $Rag1^{-/-}$ C57BL/6 mice (105 ± 5 (\pm s.e.m.) days) differed strikingly from that in MCA-treated wild-type mice rendered immunodeficient at day 200 using the anti-CD4/-CD8/-IFN γ mixture (25 ± 6 days) (Fig. 2e). Similar findings were made using 129/SvEv strain mice: 186 ± 7 days for MCA-treated $Rag2^{-/-}$ mice versus 38 ± 6 days for MCA-treated wild-type mice immunodepleted at day 230 (Fig. 2f). Taken together, these data argue strongly against continuous *de novo* transformation

as the mechanism underlying the late tumor outgrowth in wild-type mice after immunodepletion.

We therefore considered the possibility that at least some MCA-treated wild-type mice that remained free of progressively growing tumours harboured fully transformed sarcoma cells, the outgrowth of which was immunologically restrained. In the MCA model,

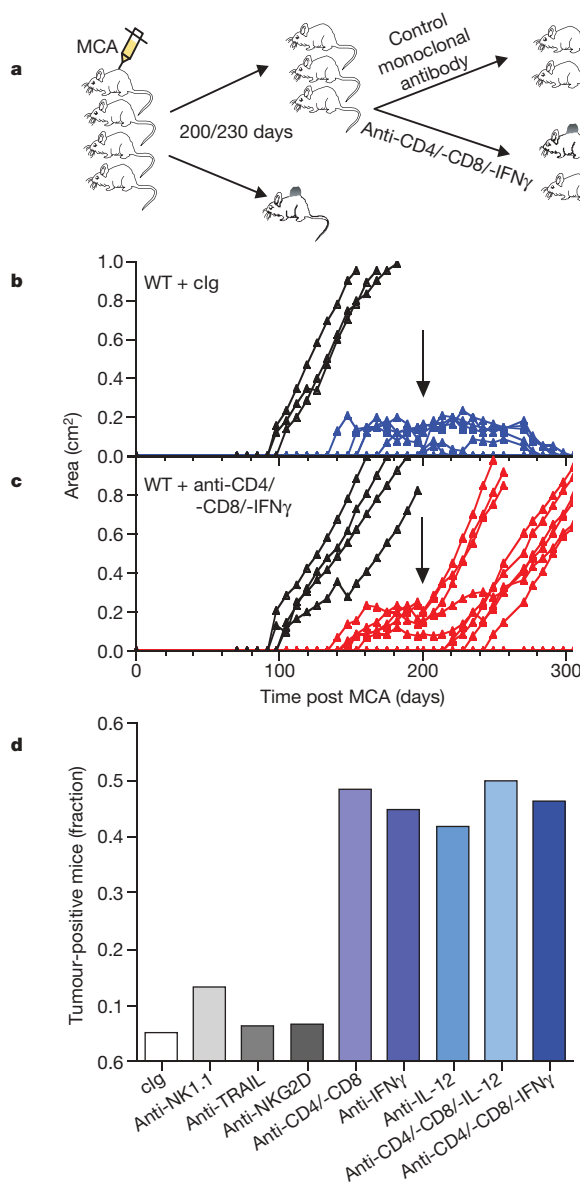


Figure 1 | The adaptive immune system promotes an equilibrium state in primary MCA-induced sarcomas. **a**, Protocol to test for the existence of an equilibrium state during MCA tumorigenesis. Details are described in the text and Methods. **b**, Sixteen wild-type (WT) C57BL/6 mice were treated with 25 μ g MCA. At 200 days, the 13 tumour-free mice were treated weekly with control immunoglobulin (cIg) and monitored for the appearance of late-forming tumours. **c**, Nineteen WT C57BL/6 mice were treated as in **b**. At 200 days, the 15 tumour-free mice were injected weekly with anti-CD4/-CD8/-IFN γ ($P = 0.0008$ Fisher's exact test between anti-CD4/-CD8/-IFN γ and cIg). **d**, Fraction of WT C57BL/6 mice treated with 25 μ g MCA that develop late-forming tumours after treatment with the following monoclonal antibodies starting at day 200: cIg ($n = 76$), anti-NK1.1 ($n = 30$), anti-TRAIL ($n = 16$), anti-NKG2D ($n = 15$), anti-CD4/-CD8 ($n = 29$), anti-IFN γ ($n = 29$), anti-IL-12p40 ($n = 12$), anti-CD4/-CD8/-IL-12 ($n = 12$) and anti-CD4/-CD8/-IFN γ ($n = 13$).

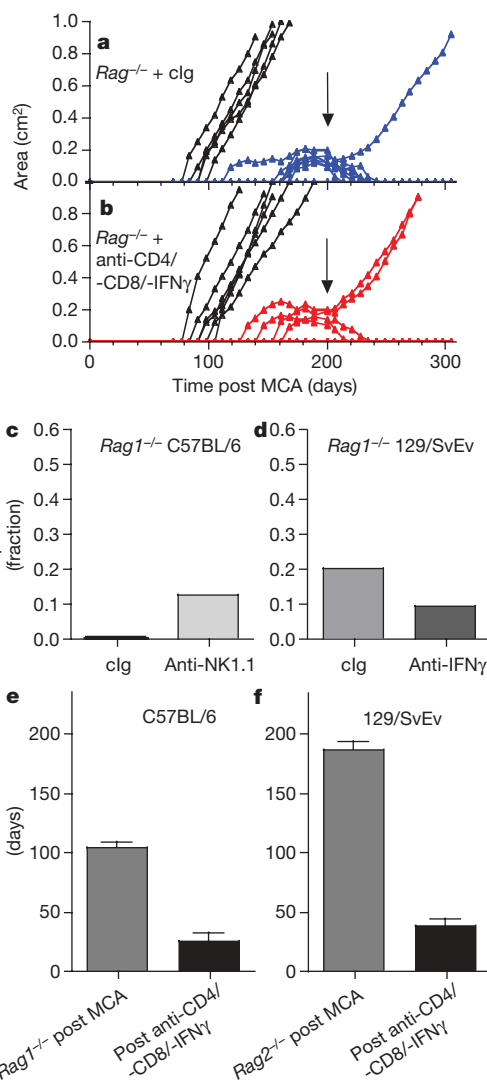


Figure 2 | Antibody-induced equilibrium disruption does not occur as a result of prolonged *de novo* transformation. $Rag1^{-/-}$ C57BL/6 mice were injected with 5 μ g MCA and, at day 200, 21 tumour-free mice were treated weekly with cIg (**a**), whereas 20 other tumour-free mice were treated with anti-CD4/-CD8/-IFN γ (**b**). Mice were monitored for the appearance of late-forming sarcomas. The difference between anti-CD4/-CD8/-IFN γ and cIg groups is not significant ($P = 0.616$, Fisher's exact test). **c**, Thirty-two tumour-free $Rag1^{-/-}$ C57BL/6 mice, injected initially with 5 μ g MCA, were treated weekly with either cIg ($n = 16$) or anti-NK1.1 ($n = 16$), starting at day 200 and monitored for sarcoma appearance. The difference between anti-NK1.1 and cIg groups is not significant ($P = 0.484$, Fisher's exact test). **d**, Twenty-seven tumour-free $Rag2^{-/-}$ 129/SvEv mice, injected initially with 25 μ g MCA, were treated weekly with either cIg ($n = 15$) or anti-IFN γ ($n = 12$) starting at day 230 and followed for sarcoma appearance. The difference between anti-IFN γ and cIg groups is not significant ($P = 0.605$, Fisher's exact test). **e**, Mean time to tumour formation \pm s.e.m. for $Rag1^{-/-}$ C57BL/6 mice treated with 25 μ g MCA ($n = 11$) or tumour-free WT C57BL/6 mice treated initially with 25 μ g MCA and subsequently with anti-CD4/-CD8/-IFN γ starting at day 200 ($n = 8$) ($P = 0.0003$, Mann-Whitney rank sum test). **f**, Mean time to tumour formation \pm s.e.m. for $Rag2^{-/-}$ 129/SvEv mice treated with 25 μ g MCA ($n = 30$) or tumour-free WT 129/SvEv mice treated initially with 25 μ g MCA and subsequently with anti-CD4/-CD8/-IFN γ starting at day 230 ($n = 30$) ($P < 0.0001$, Mann-Whitney rank sum test).

cellular transformation and tumour development occur exclusively at the site of carcinogen injection. Examination of the injection site in 129/SvEv mice treated with 25 μ g MCA revealed the presence of small 2–8 mm masses that became palpable within 150 days of MCA injection but did not change in size during an additional 150 days. When sectioned and analysed by haematoxylin and eosin staining, 47% (8/17) of the stable masses contained clusters of large cells showing variable degrees of atypia, with enlarged vesicular nuclei, prominent nucleoli and heterogeneous morphologies (Fig. 3a, c), and which also expressed vimentin (Fig. 3c, inset). These features mirrored the immuno-cytological spectrum observed in progressively growing primary MCA sarcomas (Fig. 3b, d). The remaining masses from MCA-treated wild-type mice showed variable degrees of inflammation associated with areas of dense fibrosis and lipid droplets (the oil diluent of the MCA) often surrounded by multinucleated giant cells (Supplementary Fig. 3a, c). The MCA injection site in *Rag2*^{-/-} mice also contained fibrotic areas and lipid droplets but lacked both atypical cells and multinucleated giant cells (Supplementary Fig. 3b, d). Immunohistochemical staining of stable masses containing atypical cells from MCA-treated wild-type mice and growing primary MCA sarcomas from wild-type mice revealed the presence of CD3⁺ T cells, B220⁺ cells and F4/80⁺ mononuclear phagocytes infiltrating into regions containing atypical cells (Supplementary Fig. 4). When stable masses containing atypical cells were transiently cultured, a population of atypical fibroblast-like cells grew out that formed progressively growing tumours when transplanted into immunodeficient *Rag2*^{-/-} mice (Fig. 3i). In contrast, normal skin fibroblasts from either MCA-treated mice (Fig. 3i) or untreated

wild-type mice (data not shown) did not form tumours in *Rag2*^{-/-} mice. Thus, atypical cells from stable masses that formed in MCA-treated, wild-type mice were fully transformed.

These results contradict an earlier report stating that MCA treatment of wild-type mice leads to encapsulation of MCA crystals without apparent cellular transformation¹¹. However, the former study failed to characterize 'cells with altered morphology' that were noted in the MCA granulomas and that may have been identical to the transformed atypical cells that we describe herein. Moreover, MCA crystals are not observed in our mice, indicating that the MCA might not have been fully dissolved in the earlier study and therefore may have achieved extremely high *in vivo* concentrations resulting in rapid cellular transformation and tumour outgrowth, thereby obscuring detection of transformed cells, the growth of which was under immunologic restraint.

The paradox that stable masses from our MCA-treated immunocompetent mice often contained transformed cells but did not increase in size *in vivo* suggested that net tumour cell expansion was being immunologically restrained. This possibility was explored by comparing sections of stable masses containing atypical cells or progressively growing primary sarcomas for expression of markers of proliferation (Ki-67)^{12,13} (Fig. 3e, f) or apoptosis (TUNEL) (Fig. 3g, h). Analysis of 10 progressively growing MCA sarcomas revealed strong nuclear reactivity for Ki-67 in a high proportion of tumour cells (mean proliferation index = 40%) (Fig. 3f, j), and detectable levels of cellular apoptosis (Fig. 3h). This result is consistent with the high histological grade of these lesions shown by cellular pleiomorphism, numerous visible mitoses and abundant necrosis. In contrast, only a limited percentage of atypical cells in nine stable masses stained positively for Ki-67 (mean proliferation index = 18%) (Fig. 3e, j) and these samples also showed 1.7-fold more TUNEL staining than growing sarcomas (Fig. 3g). Of note, both sample sets contained populations of proliferating, Ki-67⁺ leukocytes. Thus, stable masses are characterized by a combination of increased apoptosis and decreased tumour cell proliferation.

The concept that the immune system could maintain tumours in a 'dormant' state originally stemmed from experiments involving transplantation of certain tumour cell lines into pre-immunized mice^{14,15}. In the current study, we show that immunity can restrain the outgrowth of occult tumour cells during primary tumour induction in naive mice and additionally demonstrate that the lack of net expansion of 'dormant' tumour cells in MCA-treated immunocompetent mice is most probably the result of a continuously ongoing combination of cytostatic and cytolytic immune effects. Therefore the term 'equilibrium' best describes the persistence of cancer that we observe in wild-type mice exposed to low-dose MCA.

Throughout these experiments we noticed that a small percentage of low-dose MCA-challenged wild-type mice, either left untreated, or treated with control monoclonal antibody, went on to develop late-appearing sarcomas (Fig. 1d, Supplementary Fig. 2). We therefore asked whether sarcoma cells that spontaneously escape equilibrium could be distinguished on the basis of immunogenicity from those remaining in equilibrium. As controls, we also confirmed the immunogenicities of MCA sarcoma cell lines generated in the current study that grew out with normal kinetics from MCA-treated wild-type mice (edited tumour cells) and *Rag2*^{-/-} mice (unedited tumour cells). In agreement with previous reports¹, edited MCA-sarcoma cell lines from wild-type mice displayed dampened immunogenicity because they formed progressively growing tumours in either *Rag2*^{-/-} or naive wild-type recipients (Fig. 4a, b, respectively), and unedited MCA sarcoma cell lines from *Rag2*^{-/-} mice were highly immunogenic because, although they grew in *Rag2*^{-/-} recipients (Fig. 4a), 40% were rejected in naive wild-type mice (Fig. 4b). Notably, tumour cells that had been maintained in equilibrium (that is, derived either directly from stable masses or from MCA-treated, wild-type mice that had undergone anti-CD4/-CD8/-IFN γ -mediated equilibrium disruption) formed progressively growing tumours in

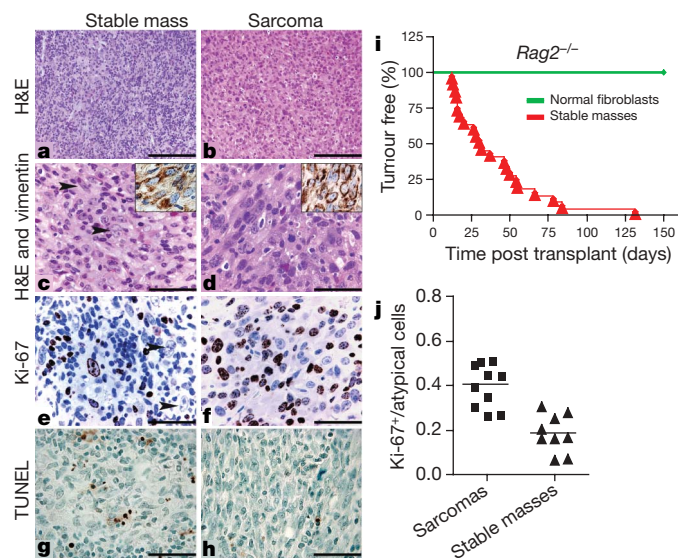


Figure 3 | Demonstration of occult cancer in immunocompetent mice in the equilibrium phase of cancer immunoediting. Sections from representative stable masses isolated from WT 129/SvEv mice treated with 25 μ g MCA (a, c, e, g) and progressively growing sarcomas from MCA-treated WT 129/SvEv mice (b, d, f, h) stained for haematoxylin and eosin (H&E) (a, b, c, d), vimentin (inset in c and d), the proliferation marker Ki-67 (e, f) and the apoptosis marker TUNEL (g, h). In c and e, arrowheads indicate atypical cells. Magnification: 200 \times ; scale bars, 100 μ m (a and b), 600 \times ; scale bars, 33 μ m (c–f) and 1,000 \times ; scale bars, 20 μ m (g and h). i, Fraction of tumour-free *Rag2*^{-/-} 129/SvEv mice following subcutaneous injection with 10⁶ normal skin fibroblasts (n = 1, diamonds) or 10⁶ cells isolated from long-term stable masses from MCA-treated WT 129/SvEv mice (n = 5, triangles). Each cell line was injected into four *Rag2*^{-/-} recipients. j, Proportion of atypical cells staining positive for Ki-67 in formalin-fixed tissue from growing primary MCA-sarcomas from WT mice (n = 10, squares) or stable masses from MCA-treated WT 129/SvEv mice (n = 9, triangles) (P = 0.0003 Mann–Whitney ranked sum test, difference between proliferation indices of tumour cells in stable masses versus growing primary MCA sarcomas).

Rag2^{-/-} mice (Fig. 3i and 4a, respectively) but were highly immunogenic in wild-type mice because 51% and 31%, respectively, were rejected (Fig. 4b). In contrast, cell lines produced from late-forming sarcomas, which grew out spontaneously from mice in equilibrium, formed tumours when transplanted into either *Rag2*^{-/-} mice or naive wild-type recipients (Fig. 4a, b, respectively). Thus, tumour cells held in equilibrium have an unedited phenotype, whereas those that spontaneously progress to the escape phase are edited. We therefore conclude that at least some tumour cell editing occurs at the temporal interface between equilibrium and escape.

In summary, in addition to the immune system's capacity to destroy and shape cancer, this study shows that immunity can also control cancer for long time periods by a process called equilibrium. We show that equilibrium is indeed a component of cancer immunoediting because tumour cells in equilibrium are highly immunogenic (unedited), whereas those spontaneously exiting equilibrium that become growing tumours have attenuated immunogenicity (edited)—results that place this process temporally between elimination and escape. We demonstrate that elimination and equilibrium can be mechanistically distinguished because, although the former requires the actions of both innate and adaptive immunity⁸, equilibrium is maintained solely by adaptive immunity. Similarly, we show

that equilibrium and escape are distinct because, whereas equilibrium represents a time of tumour cell persistence without expansion, escape is characterized by progressive tumour growth. We do not envisage that every tumour cell must pass through an elimination process before it enters equilibrium, nor do we hold that every progressively growing tumour must transit through an equilibrium process. However, we predict that many clinically apparent tumours may progress through a linear 'elimination→equilibrium→escape' continuum, and ongoing work is focused on obtaining molecular signatures of tumour cells in each step of this process. Nevertheless, our results confirm an important, but heretofore untested, prediction of the cancer immunoediting hypothesis—the existence of an equilibrium state—thereby providing additional support to the central premise of cancer immunoediting: that immunity can influence cancer development both quantitatively and qualitatively. Our findings in this mouse model also have potential relevance to human cancers. First, they indicate that maintaining cancer in an equilibrium state may represent a relevant goal of cancer immunotherapy in which augmentation of adaptive tumour immunity could result in improved tumour control. Second, they provide mechanistic underpinnings for the recent findings that the quality and quantity of the immune reaction within certain tumour types (for example, colorectal and ovarian cancers) are reliable prognostic indicators of cancer patient survival^{16–18}. Third, they explain how occult cancer can be transplanted from organ donor to recipient¹⁰, because tumour cells held in equilibrium in the donor may grow in a recipient who is naive to the antigens of the transplanted tumour cells and is immunosuppressed. Fourth, they provide a mechanism that can explain the presence of occult tumour cells in organs—for example, the prostate—of individuals lacking clinical symptoms of disease^{19–22}. Finally, our study raises the possibility that at least some of the proposed tumour-promoting actions of chronic inflammation may be a result of interfering with adaptive immunity's capacity to hold occult cancers in equilibrium^{23,24}. Our results thus provide a foundation for future work to define the molecular mechanisms by which adaptive immunity maintains cancer in an equilibrium/dormant/persistent state, perhaps paving the way for development of new therapeutic modalities to convert cancer into a controllable chronic disease.

METHODS SUMMARY

Wild-type or *Rag*-null mice were injected subcutaneously with low doses of MCA (Sigma Fine Chemicals), as described^{1,25}, and monitored for tumour development. After 200 days (C57BL/6) or 230 days (129/SvEv), MCA-treated mice bearing progressively growing tumours were removed from the experiment and the remaining mice were treated weekly with either control immunoglobulin or monoclonal antibodies that deplete or block specific immune components and were monitored for tumour development for the next 100 days. Tumour transplantation experiments were performed, as described previously^{1,25}. For morphologic and immunohistologic evaluation of primary growing sarcomas and stable masses, sections of formalin-fixed, paraffin-embedded tissue samples were evaluated after staining with haematoxylin and eosin or monoclonal antibodies specific for either cell surface markers or the Ki-67 protein. Apoptosis was determined by TUNEL staining. The proliferation index is defined as the percentage of large, atypical, cells showing enlarged nuclei with prominent nucleoli that stained positively for Ki-67 in multiple high power (600×) fields.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 June; accepted 24 September 2007.

Published online 18 November 2007.

- Shankaran, V. *et al.* IFN γ and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature* 410, 1107–1111 (2001).
- Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J. & Schreiber, R. D. Cancer immunoediting: from immunosurveillance to tumor escape. *Nature Immunol.* 3, 991–998 (2002).
- Street, S. E., Trapani, J. A., MacGregor, D. & Smyth, M. J. Suppression of lymphoma and epithelial malignancies effected by interferon γ . *J. Exp. Med.* 196, 129–134 (2002).

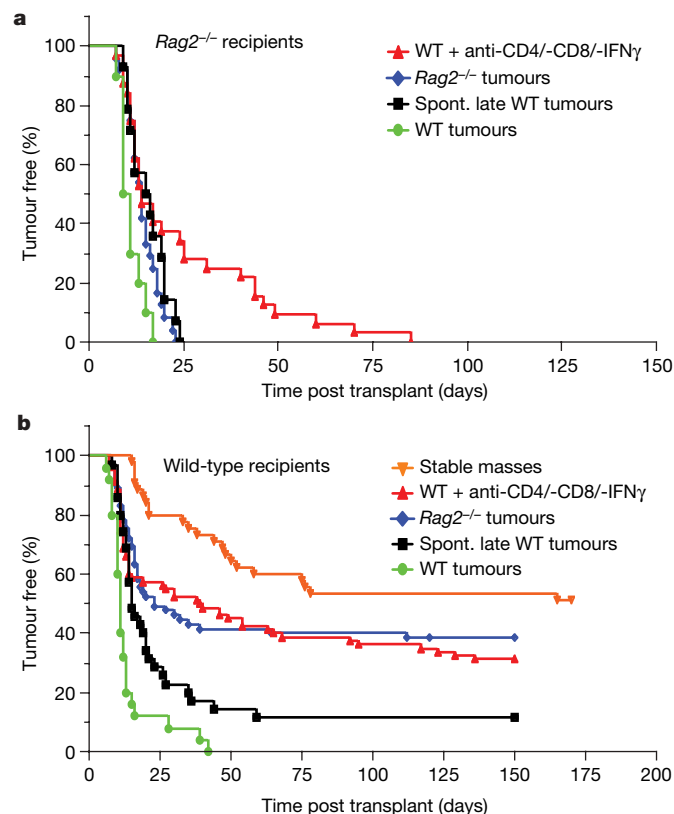


Figure 4 | Sarcoma cells in equilibrium show high immunogenicity, whereas those spontaneously exiting equilibrium have attenuated immunogenicity. **a**, Kaplan–Meier survival analysis of *Rag2*^{-/-} 129/SvEv mice injected with sarcoma cell lines from classical progressively growing MCA-sarcomas isolated from WT ($n = 5$, circles) or *Rag2*^{-/-} 129/SvEv mice ($n = 13$, diamonds) or isolated from tumours forming in MCA-treated WT 129/SvEv mice after anti-CD4/-CD8/-IFN γ -mediated equilibrium disruption ($n = 16$, triangles) or isolated from late-appearing tumours from MCA-treated WT 129/SvEv mice that spontaneously progress from equilibrium to escape ($n = 7$, squares). Each cell line was injected at 10^6 cells per mouse into two mice each. **b**, Kaplan–Meier survival analysis as in **a** except that 5-member groups of naive WT 129/SvEv mice were used as the recipients for each cell line. Also tested were sarcoma cell lines grown directly out of stable masses from MCA-treated WT mice in equilibrium ($n = 5$, inverted triangles).

4. Smyth, M. J. *et al.* NKG2D recognition and perforin effector function mediate effective cytokine immunotherapy of cancer. *J. Exp. Med.* **200**, 1325–1335 (2004).
5. Dunn, G. P., Old, L. J. & Schreiber, R. D. The immunobiology of cancer immunosurveillance and immunoediting. *Immunity* **21**, 137–148 (2004).
6. Dunn, G. P. *et al.* A critical function for type I interferons in cancer immunoediting. *Nature Immunol.* **6**, 722–729 (2005).
7. Dunn, G. P., Old, L. J. & Schreiber, R. D. The three Es of cancer immunoediting. *Annu. Rev. Immunol.* **22**, 329–360 (2004).
8. Smyth, M. J., Dunn, G. P. & Schreiber, R. D. Cancer immunosurveillance and immunoediting: the roles of immunity in suppressing tumor development and shaping tumor immunogenicity. *Adv. Immunol.* **90**, 1–50 (2006).
9. Dunn, G. P., Koebel, C. M. & Schreiber, R. D. Interferons, immunity and cancer immunoediting. *Nature Rev. Immunol.* **6**, 836–848 (2006).
10. MacKie, R. M., Reid, R. & Junor, B. Fatal melanoma transferred in a donated kidney 16 years after melanoma surgery. *N. Engl. J. Med.* **348**, 567–568 (2003).
11. Qin, Z., Kim, H. J., Hemme, J. & Blankenstein, T. Inhibition of methylcholanthrene-induced carcinogenesis by an interferon γ receptor-dependent foreign body reaction. *J. Exp. Med.* **195**, 1479–1490 (2002).
12. Brown, D. C. & Gatter, K. C. Ki67 protein: the immaculate deception? *Histopathology* **40**, 2–11 (2002).
13. Bullwinkel, J. *et al.* Ki-67 protein is associated with ribosomal RNA transcription in quiescent and proliferating cells. *J. Cell. Physiol.* **206**, 624–635 (2006).
14. Weinhold, K. J., Goldstein, L. T. & Wheelock, E. F. Tumour-dormant states established with L5178Y lymphoma cells in immunised syngeneic murine hosts. *Nature* **270**, 59–61 (1977).
15. Siu, H., Vitetta, E. S., May, R. D. & Uhr, J. W. Tumor dormancy. I. Regression of BCL1 tumor and induction of a dormant tumor state in mice chimeric at the major histocompatibility complex. *J. Immunol.* **137**, 1376–1382 (1986).
16. Galon, J. *et al.* Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* **313**, 1960–1964 (2006).
17. Ohtani, H. Focus on TILs: Prognostic significance of tumor infiltrating lymphocytes in human colorectal cancer. *Cancer Immun.* **7**, 4–12 (2007).
18. Sato, E. *et al.* Intraepithelial CD8⁺ tumor-infiltrating lymphocytes and a high CD8⁺/regulatory T cell ratio are associated with favorable prognosis in ovarian cancer. *Proc. Natl Acad. Sci. USA* **102**, 18538–18543 (2005).
19. Sakr, W. A. & Partin, A. W. Histological markers of risk and the role of high-grade prostatic intraepithelial neoplasia. *Urology* **57**, 115–120 (2001).
20. Welch, H. G. & Black, W. C. Using autopsy series to estimate the disease “reservoir” for ductal carcinoma *in situ* of the breast: how much more breast cancer can we find? *Ann. Intern. Med.* **127**, 1023–1028 (1997).
21. Mutter, G. L. *et al.* Molecular identification of latent precancers in histologically normal endometrium. *Cancer Res.* **61**, 4311–4314 (2001).
22. Bach, P. B. *et al.* Computed tomography screening and lung cancer outcomes. *J. Am. Med. Assoc.* **297**, 953–961 (2007).
23. Balkwill, F. & Coussens, L. M. Cancer: an inflammatory link. *Nature* **431**, 405–406 (2004).
24. Lin, W. W. & Karin, M. A cytokine-mediated link between innate immunity, inflammation, and cancer. *J. Clin. Invest.* **117**, 1175–1183 (2007).
25. Smyth, M. J. *et al.* Differential tumor surveillance by natural killer (NK) and NKT cells. *J. Exp. Med.* **191**, 661–668 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors are grateful for the advice of E. Unanue, G. Dunn, H. Virgin, P. Allen, M. Colonna, J. Trapani, R. Uppaluri, J. Bui and all members of the Schreiber laboratory during the preparation of this manuscript. We also greatly appreciate the technical assistance of C. Arthur, M. White, J. Archambault and J. Sharkey. This work was supported by grants to R.D.S. from the National Cancer Institute, the Ludwig Institute for Cancer Research, the Cancer Research Institute and Atlantic Philanthropies, and to M.J.S. from the National Health and Medical Research Council of Australia for Fellowship and Program Grant Support. C.M.K. was supported by a pre-doctoral training grant from the Cancer Research Institute. J.B.S. was supported by an Australian Postgraduate Research Award.

Author Contributions The work in this paper reflects an equal contribution from the M.J.S. and R.D.S. laboratories. C.M.K., M.J.S. and R.D.S. were involved in all aspects of experimental work, project planning and data analysis. W.V. and S.J.R. were responsible for performing and interpreting the histological analyses. L.J.O. and J.B.S. participated in project planning and N.Z. was involved in the experimental work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.D.S. (schreiber@immunology.wustl.edu) or M.J.S. (mark.smyth@petermac.org).

METHODS

Mice. Wild-type C57BL/6 mice were purchased from the Walter and Eliza Hall Institute of Medical Research and *Rag1*^{-/-} C57BL/6 mice were bred at the Peter MacCallum Cancer Centre. Wild-type and *Rag2*^{-/-} 129/SvEv mice were purchased from Taconic Farms and entered into the experiments at 8 to 12 weeks old. Mice were housed according to the American Association for Laboratory Animal Science conditions in specific pathogen-free facilities at the Washington University School of Medicine, St. Louis, Missouri and the Peter MacCallum Cancer Centre, East Melbourne, Victoria, Australia.

Reagents. Anti-CD4 (GK1.5)²⁶, anti-CD8 (YTS-169.4)²⁷, anti-IFN γ (H22)²⁸, anti-NKG2D (C7)²⁹, anti-NK1.1 (PK136)³⁰, anti-TRAIL (N2B2)³¹, anti-IL-12p40 (C17.8)³² and control immunoglobulin (the PIP monoclonal antibody specific for glutathione S-transferase³³) were generated from spent culture supernatant or hybridoma ascites and purified in aggregate-, endotoxin-, and protein A-free form. Antibodies were injected at an initial dose of 250–750 μ g each with a weekly (CD4, CD8, NK1.1, IFN γ) or biweekly (NKG2D, TRAIL, IL-12p40) maintenance dose of 250 μ g of each.

MCA tumour induction. Groups of wild-type and *Rag1*^{-/-} C57BL/6 mice were injected subcutaneously in the hind flank with 5 or 25 μ g of MCA in 0.10 ml of corn oil, as described³⁴. Mice were monitored every 7 days for tumour development from 70 days after MCA treatment. Tumours >0.5 cm² in area and demonstrating progressive growth were recorded as 'tumour positive'. Wild-type and *Rag2*^{-/-} 129/SvEv mice were injected subcutaneously in the flank with 25 μ g MCA (Sigma Fine Chemicals) dissolved in 0.15 ml of peanut oil (with gentle heating), as described³⁵. Mice were monitored every 7 days for tumour development from 90 days after MCA treatment. 129/SvEv mice were considered tumour positive when masses reached an average diameter of 9 mm and continued to grow progressively.

Cell lines. Cell lines were created by mechanical disruption of tissue, followed by a 1 h treatment with collagenase Type IA (0.5 mg ml⁻¹) and culture in endotoxin-low RPMI medium (<0.01 EU ml⁻¹) supplemented with 10% fetal calf serum (0.25 EU ml⁻¹).

Tumour transplantation. Before use, a vial of frozen sarcoma cells was thawed and cultured *in vitro* in RPMI medium supplemented with 10% fetal calf serum for two passages. Tumour cells were collected by incubation in 0.25% trypsin, washed two times in RPMI medium and one time in endotoxin-free PBS and then injected subcutaneously in a volume of 0.15 ml PBS into the shaved flanks of mice. All injected cell lines were more than 90% viable, as determined by trypan blue exclusion. Tumour growth was monitored by measuring two perpendicular diameters.

Histology, proliferation index and immunohistochemistry. Sections were obtained from formalin-fixed, paraffin-embedded tissue samples. For morphological evaluation, sections were stained in haematoxylin and eosin. The proliferation index was calculated from tissue sections stained using an indirect

immunoperoxidase technique that employed as primary antibody the clone Tec-3 (DakoCytomation, 1:25 dilution), which is specific for the Ki-67 protein, a marker of proliferating cells, and a biotinylated rabbit anti-rat IgG secondary antibody (Vector Laboratories, 1:200 dilution). At least 200 atypical/neoplastic cells were counted and scored per sample. All other immunohistochemical stains were performed on formalin-fixed, paraffin-embedded 4 μ m tissue sections using standard techniques. The following primary antibodies were used: anti-vimentin (Abcam, ab28028, 1:100 dilution following proteinase K digestion for 15 min.), anti-CD3 (Cell Marque, CMC363, 1:1,500 dilution), anti-CD45R/B220 (BD Pharmingen, no. 550286, 1:200 dilution), and anti-F4/80 antigen (Serotec, MCA497, 1:5,000 dilution following proteinase K digestion for 10 min.). Chromogenic terminal-deoxynucleotidyl-transferase-mediated nick-end labelling (TUNEL) staining with methyl green counterstaining was done as per the manufacturer's instructions (ApopTag Peroxidase In situ Kit, no. S7100, Chemicon International).

Statistical analysis. Fisher's exact test was used to determine the significance of the association between two variables (tumour development and antibody treatment) in a 2 \times 2 contingency table. The Mann-Whitney rank sum test was used to assess whether two samples of observations (time to tumour formation and proliferation index) come from the same distribution, without assuming equal variances between the two populations.

26. Dialynas, D. P. *et al.* Characterization of the murine antigenic determinant, designated L3T4a, recognized by monoclonal antibody GK1.5: expression of L3T4a by functional T cell clones appears to correlate primarily with class II MHC antigen-reactivity. *Immunol. Rev.* **74**, 29–56 (1983).
27. Cobbold, S. P., Jayasuriya, A., Nash, A., Prospero, T. D. & Waldmann, H. Therapy with monoclonal antibodies by elimination of T-cell subsets *in vivo*. *Nature* **312**, 548–551 (1984).
28. Schreiber, R. D., Hicks, L. J., Celada, A., Buchmeier, N. A. & Gray, P. W. Monoclonal antibodies to murine γ -interferon which differentially modulate macrophage activation and antiviral activity. *J. Immunol.* **134**, 1609–1618 (1985).
29. Ho, E. L. *et al.* Costimulation of multiple NK cell activation receptors by NKG2D. *J. Immunol.* **169**, 3667–3675 (2002).
30. Koo, G. C. & Peppard, J. R. Establishment of monoclonal anti-Nk-1.1 antibody. *Hybridoma* **3**, 301–303 (1984).
31. Kayagaki, N. *et al.* Expression and function of TNF-related apoptosis-inducing ligand on murine activated NK cells. *J. Immunol.* **163**, 1906–1913 (1999).
32. Wysocka, M. *et al.* Interleukin-12 is required for interferon- γ production and lethality in lipopolysaccharide-induced shock in mice. *Eur. J. Immunol.* **25**, 672–676 (1995).
33. Dunn, G. P. *et al.* A critical function for type I interferons in cancer immunoeediting. *Nature Immunol.* **6**, 722–729 (2005).
34. Smyth, M. J. *et al.* Differential tumor surveillance by natural killer (NK) and NKT cells. *J. Exp. Med.* **4**, 661–668 (2000).
35. Shankaran, V. *et al.* IFN γ and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature* **410**, 1107–1111 (2001).

LETTERS

The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA

Jafar Sharif^{1,2,3*}, Masahiro Muto^{4*}, Shin-ichiro Takebayashi^{6*}, Isao Suetake⁷, Akihiro Iwamatsu⁸, Takaho A. Endo⁵, Jun Shinga⁴, Yoko Mizutani-Koseki⁴, Tetsuro Toyoda⁵, Kunihiro Okamura², Shoji Tajima⁷, Kohzoh Mitsuya¹, Masaki Okano⁶ & Haruhiko Koseki⁴

DNA methyltransferase (cytosine-5) 1 (Dnmt1) is the principal enzyme responsible for maintenance of CpG methylation and is essential for the regulation of gene expression, silencing of parasitic DNA elements, genomic imprinting and embryogenesis^{1–4}. Dnmt1 is needed in S phase to methylate newly replicated CpGs occurring opposite methylated ones on the mother strand of the DNA, which is essential for the epigenetic inheritance of methylation patterns in the genome. Despite an intrinsic affinity of Dnmt1 for such hemi-methylated DNA⁵, the molecular mechanisms that ensure the correct loading of Dnmt1 onto newly replicated DNA *in vivo* are not understood. The Np95 (also known as Uhrf1 and ICBP90) protein binds methylated CpG through its SET and RING finger-associated (SRA) domain⁶. Here we show that localization of mouse Np95 to replicating heterochromatin is dependent on the presence of hemi-methylated DNA. Np95 forms complexes with Dnmt1 and mediates the loading of Dnmt1 to replicating heterochromatic regions. By using Np95-deficient embryonic stem cells and embryos, we show that Np95 is essential *in vivo* to maintain global and local DNA methylation and to repress transcription of retrotransposons and imprinted genes. The link between hemi-methylated DNA, Np95 and Dnmt1 thus establishes key steps of the mechanism for epigenetic inheritance of DNA methylation.

Methylation inheritance is the process of copying pre-existing methylation patterns onto the new DNA strand after DNA replication⁷. Dnmt1 prefers to methylate hemi-methylated CpG regions, which appear after the replication and repair steps, and thus has a dominant role in methylation inheritance⁸. Loading of Dnmt1 onto its targets involves proliferating cell nuclear antigen (Pcna), which promotes Dnmt1 localization to replication foci, but Pcna is not absolutely required in this process^{9,10}. Therefore, the molecular mechanisms that load Dnmt1 to the hemi-methylated CpG regions remain largely obscure. Recently, the *Arabidopsis* SRA protein VIM1 has been shown to be involved in recognition of methylated CpG and DNA methylation¹¹. A closely related human protein NP95 also binds to methylated promoters through its SRA domain, and mouse Np95 is essential for cell-cycle progression, DNA damage responses and replication of pericentromeric heterochromatin^{6,12–14}. Therefore, Np95 is a possible candidate linking Dnmt1 with hemi-methylated DNA in mammals.

To test this possibility, we first examined the localization of Np95 in embryonic stem cells (ESCs) by immunofluorescence analysis. Because Np95 is known to colocalize with replication foci in mid-to-late-S-phase fibroblasts¹⁴, we sorted ESCs into representative

cell-cycle stages and then stained them with 4,6-diamidino-2-phenylindole (DAPI), and Np95, Dnmt1 and Pcna antibodies (Fig. 1a). There was an intense accumulation of Np95 at DAPI-dense heterochromatin regions in mid-to-late-S-phase nuclei, but not in the G1 or G2/M phase. We confirmed colocalization of Np95 with Dnmt1 and Pcna in mid-to-late-S-phase ESC nuclei, an observation that prompted us to investigate whether human NP95 forms complexes with DNMT1. *In vivo* biotinylated NP95 was used in a pull-down assay¹⁵ (Supplementary Fig. 1). We tested for the presence of DNMT1 in NP95 complexes by immunoblotting, and found significant quantities (Fig. 1b) of catalytically active NP95-associated DNMT1 (ref. 16, Supplementary Fig. 2). Because DNMT1 interacts with PCNA^{9,10}, we extended the analysis to PCNA and also found PCNA in the NP95 complexes (Fig. 1b). Taken together, these results indicate that NP95 forms complexes with DNMT1 and PCNA at replicating heterochromatic regions¹⁷, where hemi-methylated DNA is generated and concurrently converted into full-methylated DNA on both strands.

We next examined whether the localization of Np95 is dependent on methylated DNA using various *Dnmt*-deficient ESCs with characteristic DNA methylation profiles. In wild-type ESCs, Np95 exhibited a focal accumulation in replicated heterochromatin in a small fraction of cells (~20%) and a diffuse localization pattern in the rest (Fig. 1c–e). A similar Np95 distribution profile was seen in *Dnmt3a*^{−/−};*Dnmt3b*^{−/−} double-knockout (DKO) ESCs (Fig. 1c–e), which retain considerable global DNA methylation¹⁸. Np95 was also able to localize to heterochromatin in *Dnmt1*^{−/−} ESCs (Fig. 1c, e), in which global DNA methylation is extensively decreased but not abolished^{18,19}. Interestingly, the percentage of *Dnmt1*^{−/−} ESCs showing Np95 heterochromatin accumulation was much higher than in wild-type and 3abDKO ESCs (Fig. 1c). In contrast, Np95 showed a diffuse localization pattern and almost no enrichment in the newly replicated heterochromatin in *Dnmt1*^{−/−};*Dnmt3a*^{−/−};*Dnmt3b*^{−/−} triple knockout (TKO) ESCs (Fig. 1c–e), in which DNA methylation is absent¹⁹. These findings support the idea that localization of Np95 to replicating heterochromatin is dependent on methylated DNA. To confirm this, we transiently expressed Dnmt3a and Dnmt3b, both of which methylate DNA at pericentric heterochromatin, as either wild-type or catalytically defective mutants in TKO ESCs and examined the heterochromatin localization of Np95 (refs 20 and 21; Fig. 1f and Supplementary Fig. 3). Dnmt3a and Dnmt3b restored the heterochromatin localization of Np95, whereas the catalytically defective mutants did not. These findings indicate that the heterochromatin

¹Tohoku University Biomedical Engineering Research Organization (TUBERO), 2-1 Seiryō-machi, Aoba-ku, Sendai 980-8575, Japan. ²Department of Obstetrics and Gynecology, Tohoku University School of Medicine, 1-1 Seiryō-machi, Aoba-ku, Sendai 980-8574, Japan. ³Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. ⁴RIKEN Research Center for Allergy and Immunology, ⁵RIKEN Genomic Sciences Center, 1-7-22 Suehiro, Tsurumi-ku, Yokohama 230-0045, Japan. ⁶RIKEN Center for Developmental Biology, 2-2-3 Minatogijima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan. ⁷Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan. ⁸Protein-Research Network, Inc., 1-13-5 Fukuura, Kanazawa-ku, Yokohama 236-0004, Japan.

*These authors contributed equally to this work.

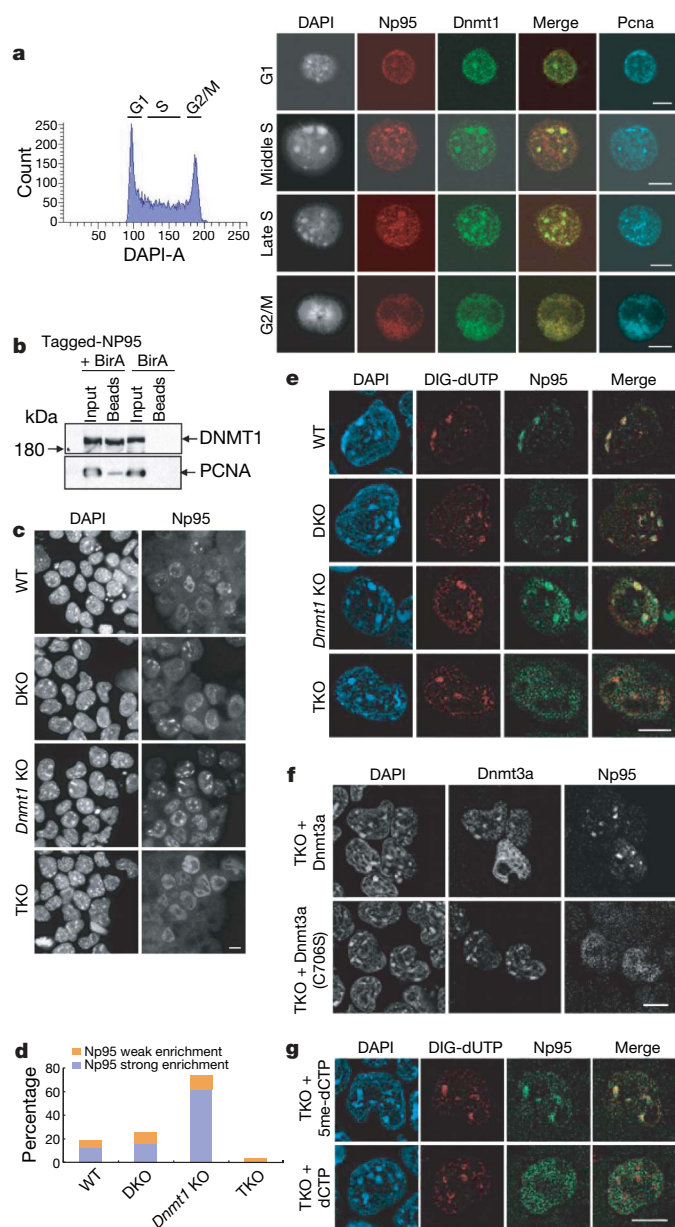


Figure 1 | Local accumulation of Np95 is dependent on hemi-methylated DNA. **a**, Representative subnuclear localization of Np95, Dnmt1 and PcnA in wild-type (WT) ESCs (E14) during cell-cycle progression (right). Merged images for Np95 and Dnmt1 are shown. Profile of DNA content in exponentially growing ESCs is shown (left). DAPI-A is an arbitrary unit that represents fluorescent intensity of cells stained by DAPI. **b**, Association of NP95 with DNMT1 and PCNA in HeLa cell nuclear extracts. HeLa cells transfected with expression vectors for BirA *Escherichia coli* biotin-protein ligase and tagged-NP95 are indicated as 'Tagged-NP95 + BirA', whereas those transfected with BirA alone are labelled 'BirA'. **c**, Immunofluorescence analysis of Np95 in wild-type, *Dnmt3a*^{-/-}; *Dnmt3b*^{-/-} DKO, *Dnmt1* knockout (KO) and *Dnmt1*^{-/-}; *Dnmt3a*^{-/-}; *Dnmt3b*^{-/-} TKO ESCs. **d**, Frequency of cells exhibiting heterochromatic localization of Np95 in respective ESCs. More than three-hundred cells were visually scored for the degree of speckled localization of Np95 for either strong or weak enrichment at the heterochromatic regions. **e**, Nuclear localization of Np95 during replication of the pericentromeric heterochromatin. Replication sites were visualized by the incorporation of digoxigenin-11 (DIG)-dUTP. **f**, Localization of Np95 in TKO ESCs, in which a wild-type or catalytic-defective (C706S) Dnmt3a was transiently expressed. **g**, Localization of Np95 in TKO ESCs, in which methylated or unmethylated dCTP was incorporated simultaneously with DIG-dUTP. Scale bars represent 10 μ m.

accumulation of Np95 during S phase is dependent on the DNA methylation mark rather than on the presence of Dnmt3 proteins.

Given that the primary Dnmt1 substrate for methylation inheritance is hemi-methylated DNA, we hypothesized that Np95 primarily recognizes hemi-methylated CpG, which is enriched in newly replicated regions and thus is distributed in the nuclei in a cell-cycle-dependent manner. Consistent with this hypothesis, Np95 accumulated in heterochromatin in most *Dnmt1*^{-/-} ESCs independently of the cell-cycle stage, whereas the heterochromatin enrichment of Np95 was barely detectable in the early S phase of the wild-type cells (Fig. 1c and Supplementary Fig. 4). This difference could be explained by increased hemi-methylated CpG in *Dnmt1*^{-/-} ESCs, as has been seen in DNMT1-defective human cancer cells²², whereas heterochromatic regions are probably symmetrically CpG-methylated in wild-type ESCs in early S phase. We thus examined colocalization of Np95 and hemi-methylated DNA by replication labelling, in which 5-methyl-dCTP is incorporated into the nascent strand of an unmethylated TKO genome²³. Np95 was prominently accumulated in heterochromatin when 5-methyl-dCTP was incorporated into TKO ESCs (Fig. 1g). These results strongly support a model in which local accumulation of Np95 is dependent on hemi-methylated DNA. Consistent with our observations, the recombinant SRA domain has been shown recently to bind hemi-methylated CpG *in vitro*²⁴.

We went on to examine the affect of Np95 loss on DNA methylation¹³ (Supplementary Fig. 5). *Np95*^{-/-} embryos show developmental arrest shortly after gastrulation and exhibit early gestational lethality in a similar manner to *Dnmt1*^{-/-} embryos. Genomic DNA isolated from *Np95*^{-/-} and wild-type ESCs or embryos was evaluated for the degree of global and local DNA methylation. *Dnmt1*^{-/-} ESCs were used as a reference in these experiments¹⁸. The level of global DNA methylation in the absence of Np95 was determined by examining the resistance of the DNA to methylation-sensitive restriction enzymes. Global CpG methylation was reduced substantially in *Np95*^{-/-} ESCs and embryos (Fig. 2a).

To refine our analysis, we next examined CpG methylation levels at heterochromatic domains. Major and minor satellites at pericentromeric and centromeric heterochromatin, respectively, are highly compacted and condensed; this is, at least in part, caused by the high levels of methylated CpG. DNA blot analyses for major and minor satellites revealed extensive hypomethylation in both *Np95*^{-/-} and *Dnmt1*^{-/-} ESCs (Fig. 2b). We further investigated CpG methylation levels by immunofluorescence using a 5-methylcytosine antibody. Intense 5-methylcytosine staining at pericentromeric regions was seen in the mitotic chromosomes and interphase nuclei of wild-type ESCs, but was significantly reduced in *Dnmt1*^{-/-} and was totally abolished in TKO cells (Fig. 2c). In *Np95*^{-/-} ESCs, 5-methylcytosine staining was reduced to a level similar to that in *Dnmt1*^{-/-} ESCs, but was less than levels in TKO cells. The methylation defect in *Np95*^{-/-} cells was complemented by expression of Myc-tagged Np95. Equivalent data were obtained for the major satellite sequences by DNA blot analyses (Supplementary Fig. 6). In summary, CpG methylation in heterochromatic regions is reduced to a similar extent in *Np95*^{-/-} and *Dnmt1*^{-/-} ESCs.

Most CpGs are methylated at retrotransposon-derived elements in euchromatic regions. Loss of DNA methylation by Dnmt1 inactivation results in derepression of silenced retrotransposons⁴. We thus examined DNA methylation at the promoter regions of the intracisternal A particle (IAP) and long interspersed nuclear element 1 (LINE-1) retrotransposons. Hypomethylation of IAP retrotransposons in *Np95*^{-/-} ESCs was demonstrated by DNA blot analysis (Fig. 2d). Moreover, hypomethylation of IAP and LINE-1 elements in *Np95*^{-/-} embryos was confirmed by bisulphite genomic sequencing (Fig. 2e). Prompted by these findings, we further investigated whether Np95 has a role in genomic imprinting, which is mediated by DNA methylation in somatic cells². CpG methylation at the imprinting control regions of imprinted *H19*, *Kcnq1ot1* (also known as *Lit1*)

and *Gtl2* genes was reduced in *Np95*^{-/-} ESCs (Fig. 2e). Taken together, these results implicate Np95 in DNA methylation at both heterochromatin and euchromatin compartments.

We went on to test whether defects in DNA methylation perturb transcriptional repression in *Np95*^{-/-} embryos. Indeed, RNAi-mediated knockdown of Np95 has been reported to increase pericentromeric transcription¹⁴. Similarly, in our analysis, IAP, LINE-1 and short interspersed nuclear element 1 (SINE-1) retrotransposons were derepressed in *Np95*^{-/-} embryos (Fig. 3a). We also examined whether parent-of-origin-specific expression at imprinted *H19* and *Kcnq1ot1* loci was retained in *Np95*^{-/-} embryos. In the wild-type embryos, maternally derived *H19* and paternally derived *Kcnq1ot1* were exclusively expressed, whereas, in *Np95*^{-/-} embryos, both alleles were expressed (Fig. 3b). This was accompanied by silencing

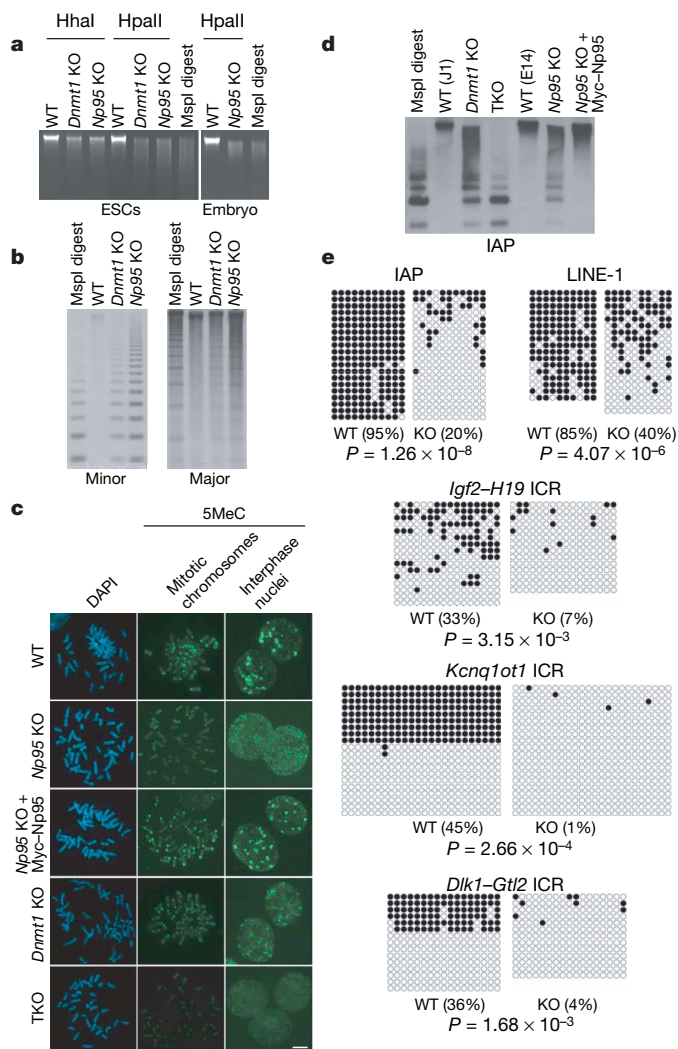


Figure 2 | Impairment of DNA methylation status on *Np95* gene inactivation. **a**, Genome-wide DNA demethylation in *Np95* KO ESCs (left) and embryos (right). **b**, DNA demethylation at the centromeric minor (left) and major (right) satellite repeats. **c**, Anti-5-methylcytosine (5MeC) immunofluorescence in wild-type (WT), *Np95* KO, *Np95* KO + Myc-Np95, *Dnmt1* KO and TKO ESCs in mitotic chromosomes and interphase nuclei. Scale bars represent 10 μ m. **d**, DNA demethylation at the IAP retrotransposons. J1 and E14 cells are the parental ESCs for *Dnmt1* and *Np95* KO, respectively. **e**, Extensive demethylation of retrotransposons and imprinting centres in *Np95* KO embryos and ESCs, respectively. Bisulphite sequencing results obtained from E9.5 mouse embryos for IAP and LINE-1, and from ESCs for imprinting control regions (ICRs) of *Igf2-H19*, *Kcnq1ot1* and *Dlk1-Gtl2* are shown. The overall percentage of methylated CpGs is indicated in parentheses. The *P* values indicate the significance of the difference between wild-type and *Np95* KO embryos.

of the adjacent imprinted transcripts within the clusters (*Igf2* and *Cdkn1c*). Therefore, Np95 is essential for transcriptional silencing of heterochromatin and retrotransposons as well as for parent-of-origin-specific expression of imprinted genes through regulation of CpG methylation status. Taken together, Dnmt1-dependent CpG methylation requires its association with Np95, indicating that Np95 is required either to stimulate the catalytic activity or to direct recruitment of Dnmt1 to its DNA targets.

To test the first possibility, we examined the expression and catalytic activity of Dnmt1 in the absence of Np95. Dnmt1 expression and catalytic activity were comparable in wild-type and *Np95*^{-/-} ESCs (Fig. 4a). We also confirmed that expression of both Dnmt3a and Dnmt3b were maintained (Supplementary Fig. 7). We then addressed the second possibility by determining whether Np95 affects Dnmt1 subnuclear localization. Dnmt1 had either a diffuse localization pattern or a focal accumulation in heterochromatin in wild-type ESCs, as shown in previous studies¹⁷, whereas in almost all *Np95*^{-/-} ESCs Dnmt1 had a diffuse localization pattern (Fig. 4b). The wild-type phenotype was restored by introduction of Myc-tagged Np95. In wild-type cells, Dnmt1 accumulated in the replicating pericentromeric heterochromatin regions in mid-S-phase, and partially retained its heterochromatin localization in late S phase. In contrast, no significant enrichment of Dnmt1 in heterochromatin was observed throughout the S phase in *Np95*^{-/-} ESCs (Fig. 4c). These results demonstrate the requirement of Np95 for proper localization of Dnmt1. Because subnuclear localization of Dnmt1 is dependent on pre-existing DNA methylation marks^{23,25} and Np95 binds methylated CpG²⁴, Np95 is shown to link Dnmt1 to methylated DNA.

Our findings extend our understanding of the molecular mechanism that ensures the fidelity and efficacy of DNA methylation. We have identified binding of the SRA-domain protein Np95 to pre-existing methylated CpG, probably in a hemi-methylated state, as an essential process in loading Dnmt1 to hemi-methylated sites where it mediates accurate and sufficient DNA methylation after

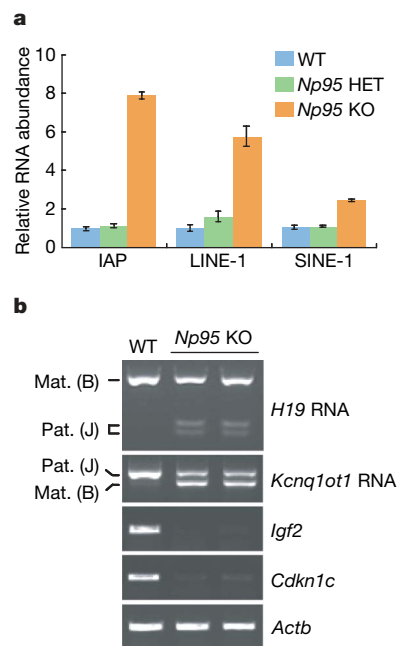


Figure 3 | Misexpression of hypomethylated genes on *Np95* gene inactivation. **a**, Transcriptional derepression of endogenous retrotransposons in *Np95* KO embryos. Error bars represent standard deviation. Total RNA derived from wild-type (WT), heterozygotes (HET) and KO embryos were used. **b**, Disruption of functional imprinting at the two imprinted gene clusters. Monoallelic expression of *H19* (paternal repression) and *Kcnq1ot1* (maternal repression) was lost completely, whereas expression of *Igf2* and *Cdkn1c* was absent in *Np95* KO embryos. B, C57BL/6 strain; J, JF1 strain; Mat., maternal; Pat., paternal; Actb, β -actin.

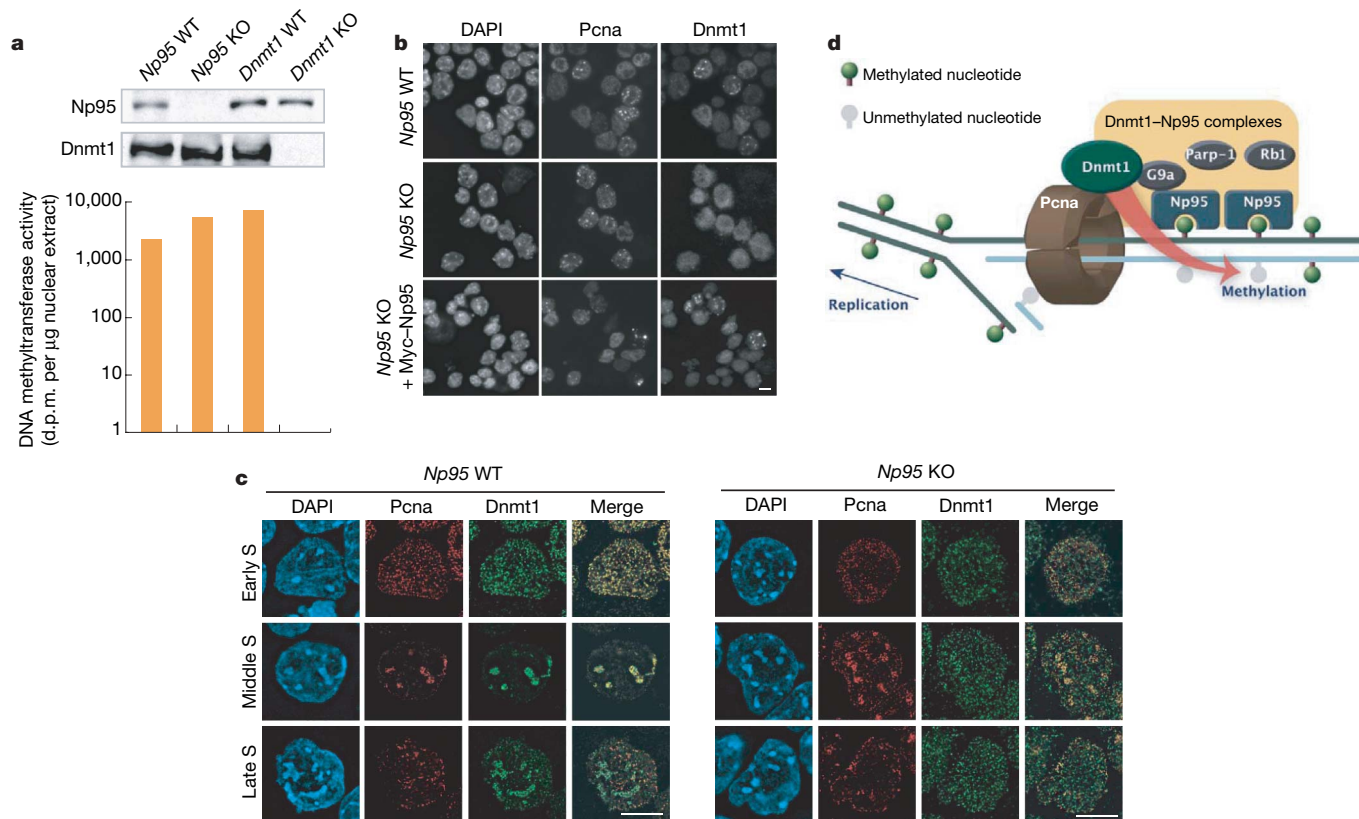


Figure 4 | Np95 is required to target Dnmt1. **a**, DNA methyltransferase activity and the expression of Dnmt1 in Np95 KO ESCs. d.p.m., decay per minute. **b**, Immunofluorescence analysis of Dnmt1 and Pcna in wild-type, Np95 KO, and Np95 KO + Myc-Np95 ESCs. **c**, Nuclear localization of Dnmt1 (green) with Pcna (red) in the early S phase, middle S phase and late S phase of wild-type or Np95 KO ESCs. The merged images at the right show

replication (Fig. 4d). Consistent with this model, the Np95 SRA domain preferentially binds to hemi-methylated CpG dinucleotides²⁴. It is noteworthy that Np95 forms complexes with nuclear proteins involved in DNA repair and chromatin modifications and/or functions^{6,14} (Supplementary Tables 1 and 2). The role of Np95 in linking Dnmt1 with DNA repair is probably represented by hypersensitivity to DNA damage observed in the absence of Np95 (ref. 13). Moreover, Np95 complexes share Pcna, Rb1, Parp1 and G9a with Dnmt1 (refs 9 and 26–28; Fig. 1b and Supplementary Table 1). This suggests the involvement of Np95 in sensing not only CpG methylation status but also other chromatin features to guarantee appropriate localization of Dnmt1 (Fig. 4d). Intriguingly, similar methylation defects are seen in cells deficient in Cxxc1, an unmethylated CpG binding protein²⁹. This suggests that more proteins recognize CpG methylation status in mediating levels of genomic methylation. Collectively, our findings provide the first evidence that the SRA-domain protein Np95 is required for a hallmark feature of the epigenome, the coordinate regulation of local and global epigenetic features.

METHODS SUMMARY

Immunofluorescence analysis. Detailed procedures for immunostaining and imaging are described in Methods. The following antibodies were used: Np95 rat monoclonal antibody¹³, Dnmt1 rabbit polyclonal antibody (Santa Cruz Biotechnology, sc-20701), Pcna mouse monoclonal antibody (Santa Cruz Biotechnology, sc-56), Alexa Fluor-488 goat anti-mouse immunoglobulin G (Molecular Probes, A11017) and Alexa Fluor-555 goat anti-rabbit immunoglobulin G (Molecular Probes, A21430).

Replication labelling. Nucleotide analogues were introduced into cells by a hypotonic shift method as described previously^{23,30}, with slight modifications as described in Methods.

overlays of Pcna and Dnmt1 staining. The relative contribution of the S-phase stage was not altered in Np95 KO ESCs (Supplementary Fig. 8). **d**, Schematic representation of molecular actions of Dnmt1-Np95 complexes at the replicating heterochromatic regions identified in this study. Pre-existing and newly synthesized DNA strands are indicated by green and blue lines, respectively. Scale bars represent 10 μ m.

Pull down of NP95 complexes from HeLa cell extracts. NP95 complementary DNA was amino-terminally tagged with an oligonucleotide encoding a 23-amino acid biotin-binding domain¹⁵ and was subcloned into pCAGGS. Procedures for cell culture and protein purification are described in Methods. **Mice.** Np95-deficient mice were generated by using R1 ESCs according to standard protocols, and backcrossed onto a C57BL/6 background for three to six generations¹³. Schematic representations of Np95 genomic organization and its targeting strategy are illustrated in Supplementary Fig. 1.

DNA methylation analysis. The degree of DNA methylation was assessed by DNA blot hybridization, bisulphite genomic sequencing and 5-methylcytosine immunostaining as described in Methods. Results of bisulphite genomic sequencing were statistically examined as described in Methods.

RNA expression analysis. PCR with reverse transcription analysis was carried out as described in the Methods. Primers used in this study are listed in Supplementary Table 3. Detailed PCR conditions used in this work are available on request.

DNA methyltransferase activity. DNA methyltransferase activity was determined as described previously¹⁶.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 July; accepted 22 October 2007.

Published online 11 November 2007.

- Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
- Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 362–365 (1993).
- Jackson-Grusby, L. *et al.* Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nature Genet.* **27**, 31–39 (2001).
- Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genet.* **20**, 116–117 (1998).

5. Fatemi, M., Hermann, A., Pradhan, S. & Jeltsch, A. The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. *J. Mol. Biol.* **309**, 1189–1199 (2001).
6. Unoki, M., Nishidate, T. & Nakamura, Y. ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. *Oncogene* **23**, 7601–7610 (2004).
7. Klose, R. J. & Bird, A. P. Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* **31**, 89–97 (2006).
8. Goll, M. G. & Bestor, T. H. Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74**, 481–514 (2005).
9. Chuang, L. S. *et al.* Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science* **277**, 1996–2000 (1997).
10. Spada, F. *et al.* DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. *J. Cell Biol.* **176**, 565–571 (2007).
11. Woo, H. R., Pontes, O., Pikaard, C. S. & Richards, E. J. VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. *Genes Dev.* **21**, 267–277 (2007).
12. Bonapace, I. M. *et al.* Np95 is regulated by E1A during mitotic reactivation of terminally differentiated cells and is essential for S phase entry. *J. Cell Biol.* **157**, 909–914 (2002).
13. Muto, M. *et al.* Targeted disruption of Np95 gene renders murine embryonic stem cells hypersensitive to DNA damaging agents and DNA replication blocks. *J. Biol. Chem.* **277**, 34549–34555 (2002).
14. Papait, R. *et al.* Np95 is implicated in pericentromeric heterochromatin replication and in major satellite silencing. *Mol. Biol. Cell* **18**, 1098–1106 (2007).
15. de Boer, E. *et al.* Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc. Natl Acad. Sci. USA* **100**, 7480–7485 (2003).
16. Suetake, I., Miyazaki, J., Murakami, C., Takeshima, H. & Tajima, S. Distinct enzymatic properties of recombinant mouse DNA methyltransferases Dnmt3a and Dnmt3b. *J. Biochem.* **133**, 737–744 (2003).
17. Leonhardt, H., Page, A. W., Weier, H. U. & Bestor, T. H. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **71**, 865–873 (1992).
18. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* **99**, 247–257 (1999).
19. Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805–814 (2006).
20. Bachman, K. E., Rountree, M. R. & Baylin, S. B. Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. *J. Biol. Chem.* **276**, 32282–32287 (2001).
21. Lin, I. G. *et al.* Murine *de novo* methyltransferase Dnmt3a demonstrates strand asymmetry and site preference in the methylation of DNA *in vitro*. *Mol. Cell. Biol.* **22**, 704–723 (2002).
22. Chen, T. *et al.* Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. *Nature Genet.* **39**, 391–396 (2007).
23. Takebayashi, S., Tamura, T., Matsuoka, C. & Okano, M. Major and essential role for DNA methylation mark in mouse embryogenesis and stable association of DNMT1 with newly replicated regions. *Mol. Cell. Biol.* **27**, 8243–8258 (2007).
24. Bostick, M. *et al.* UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317**, 1760–1764 (2007).
25. Damelin, M. & Bestor, T. H. Biological functions of DNA methyltransferase 1 require its methyltransferase activity. *Mol. Cell. Biol.* **27**, 3891–3899 (2007).
26. Pradhan, S. & Kim, G. D. The retinoblastoma gene product interacts with maintenance human DNA (cytosine-5) methyltransferase and modulates its activity. *EMBO J.* **21**, 779–788 (2002).
27. Reale, A. *et al.* Modulation of DNMT1 activity by ADP-ribose polymers. *Oncogene* **24**, 13–19 (2005).
28. Estève, P. O. *et al.* Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev.* **20**, 3089–3103 (2006).
29. Carlone, D. L. *et al.* Reduced genomic cytosine methylation and defective cellular differentiation in embryonic stem cells lacking CpG binding protein. *Mol. Cell. Biol.* **25**, 4881–4891 (2005).
30. Koberna, K. *et al.* Nuclear organization studied with the help of a hypotonic shift: its use permits hydrophilic molecules to enter into living cells. *Chromosoma* **108**, 325–335 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported in part by a grant from the Genome Network Project (to H.K.), by the 'Ground-based Research Program for Space Utilization' promoted by the Japan Space Forum (H.K.) and by a grant-in-aid for Scientific Research on Priority Areas (germ-cell development, reprogramming and epigenetics, to M.O. and K.M.) from the Ministry of Education, Culture, Sports, Science and Technology, Japan. We thank W. Reik, N. Brockdorff, P. Burrows, H. Niwa, H. Sano, J. Strouboulis and M. Vidal for critical reading and reagents.

Author Contributions J. Sharif., K.O. and K.M. performed DNA methylation and gene expression analyses; M.M., Y.M.-K. and H.K. generated, maintained and performed phenotypic analyses of knockout mice; M.M., J. Shinga. and A.I. purified the protein complexes and performed mass spectrometry analyses; S. Takebayashi and M.O. performed immunofluorescence analysis; M.M., I.S. and S. Tajima performed the DNA methylation assay; and T.A.E. and T.T. performed statistical analyses. K.M., M.O. and H.K. designed the study, wrote the paper and contributed equally as co-senior authors. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.K. (koseki@rcai.riken.jp).

METHODS

Immunofluorescence analysis. E14 ESCs grown on culture dishes were collected by trypsinization, fixed with ice-cold 70% ethanol for 30 min, stained with DAPI ($2 \mu\text{g ml}^{-1}$), and separated using a cell sorter (Becton Dickinson, FACSDiVa) according to the DNA content. These cells were immunostained as described below.

ESCs were spun onto glass slides, fixed with 4% paraformaldehyde in PBS, permeabilized with 1% Triton X-100 in PBS, incubated in blocking solution (3% BSA, 0.1% Tween 20, $4 \times \text{SSC}$), and then incubated in detection solution containing primary antibodies (1% BSA, 0.1% Tween 20, $4 \times \text{SSC}$). After three washes with $4 \times \text{SSC}$, the samples were incubated in detection solution containing the secondary antibodies. For the PcnA immunostaining, the cells were treated with 0.5% Triton X-100 in CSK buffer (100 mM NaCl, 300 mM sucrose, 10 mM PIPES, pH 6.8, 3 mM MgCl_2 and 1 mM EGTA) for 30 s at 4°C , fixed with paraformaldehyde, and then treated with methanol for 20 min at -20°C . Images were collected using a Leica DM RA2 fluorescence microscope equipped with a cooled charge-coupled device camera (C4742-95-12ER, Hamamatsu Photonics, Inc.) controlled by a Macintosh G4 computer running the software program IPLab (Signal Analytics). The images were captured at different stage positions and were processed using deconvolution software (Scientific Volume Imaging).

Replication labelling. Fifty micromolar each of DIG-dUTP (Roche Applied Science), dATP, dGTP and dCTP was added to the KHB buffer (10 mM HEPES, pH 7.4, and 30 mM KCl). The mixture was added to cells growing in a 48-well culture dish. After incubation at 37°C with 5% CO_2 for 30 s, the cells were washed and cultured in normal medium for 10 min to introduce the nucleotides into nascent DNA. DIG-labelled DNA was detected using anti-DIG antibody conjugated with rhodamine. To introduce methylated or unmethylated dCTP into the cell nucleus, 1 mM 5me-dCTP (Roche Applied Science) or dCTP was added with $50 \mu\text{M}$ DIG-dUTP to KHB buffer. After the hypotonic shift treatment, the cells were cultured for 1 h, and were harvested for immunostaining.

Pull down of NP95 complexes from HeLa cell extracts. Nuclear extracts of HeLa cells transfected with expression vectors for *E. coli* BirA ligase and/or tagged NP95 were prepared according to standard procedures. Briefly, 6×10^9 cells were harvested by centrifugation at 430g and washed twice in cold PBS. The cell pellet was resuspended in 10 mM HEPES (pH 7.9), 1.5 mM MgCl_2 , 10 mM KCl, 0.5 mM DTT and protease inhibitors (complete, EDTA-free, Roche). The cell suspension was kept on ice for 10 min and was then homogenized with a Wheaton B homogenizer. After centrifugation of the cell lysate for 10 min at 400g at 4°C and, again, at 13,600g for 20 min, the nuclear pellets were resuspended in 100 mM KCl Heng buffer (20 mM HEPES, pH 7.9, 100 mM KCl, 20% glycerol, 0.25 mM EDTA and 0.05% NP-40). Next, 2.2 M KCl Heng buffer was added to achieve a final salt concentration of 0.4 M KCl, and nuclei were incubated with rotation for 30 min at 4°C . The homogenates were ultracentrifuged at 38,000g for 1 h at 4°C , and supernatants were aliquoted, frozen in liquid

nitrogen and stored at -80°C . Nuclear extracts (7.5 mg) were adjusted to 0.1 M KCl Heng buffer with KCl-free Heng buffer, and incubated with paramagnetic streptavidin beads ($25 \mu\text{l mg}^{-1}$ protein, Dynabeads M-280), which were pre-blocked with chicken albumin for 1 h at 4°C on a rotating wheel. The bead-protein complexes were washed six times with 0.1 M KCl Heng buffer, and were resuspended in buffer containing 2.7 M glycerol, 5 mM EDTA, 0.2 mM DTT, 25 mM NaCl, 20 mM Tris-HCl, pH 7.4, and 1% Triton X-100 or SDS sample buffer for DNA methyltransferase assay or MS analysis, respectively.

DNA methylation analysis. DNA blot hybridization was performed to assess the DNA methylation status at the major and minor satellite repeats. Genomic DNA (5 μg) was digested with appropriate restriction enzymes and hybridized with PCR-generated probes. The DNA methylation status at retrotransposons and imprinting centres was determined by use of bisulphite genomic sequencing. Bisulphite treatment of the genomic DNA samples was carried out with the Qiagen EpiTect kit. The amplified fragments were cloned using the TOPO TA cloning kit (Invitrogen), and subsequently sequenced with the BigDye Terminator Cycle Sequencing system (Applied Biosystems). The probes and primers used in this study are listed in Supplementary Table 3. For immunostaining against 5-methylcytosine, ESCs were incubated in a hypotonic solution (0.075 M KCl), fixed in methanol and acetic acid (3:1), and deposited onto a glass slide. After air-drying, slides were irradiated with ultraviolet light for 8 h under a germicidal lamp. Detection of 5-methylcytosine was carried out according to standard immunostaining protocols.

Statistical methods. The non-parametric Mann-Whitney *U*-test was used to analyse the significance of the differences in DNA methylation status between wild-type and *Np95*^{-/-} embryos or ESCs. A non-parametric method was adopted because the number of methylated nucleotides did not fit a gaussian distribution. The *P* values were calculated using our in-house computer program. To perform statistical tests on imprinted genes, we slightly modified the method because we could not formally distinguish maternal and paternal alleles in these experiments. We therefore hypothesized that genomic sequences with more methylated nucleotides than the median number corresponded to imprinted genes, and performed the Mann-Whitney *U*-test between arbitrarily determined imprinted alleles of wild type and knockout.

RNA expression analysis. Fluorescence real-time PCR analysis was performed on a LightCycler instrument using the SYBR Green detection system according to the protocol provided by the manufacturer (Roche). Each experiment was conducted independently in triplicate to generate a mean and standard deviation. Data were normalized to the level of β -actin gene expression in each individual sample. Monoallelic expression of *H19* (paternal repression) and *Kcnq1ot1* (maternal repression) RNA was investigated using single nucleotide polymorphisms that distinguish the alleles from JF1 and C57BL/6 strains. Total RNA was collected from the offspring of a C57BL/6 \times (C57BL/6 \times JF1) cross in which the JF1 distal chromosome 7 was paternally inherited. Amplified fragments that included the single nucleotide polymorphisms were analysed by use of restriction-fragment length polymorphisms.

A hierarchy of timescales in protein dynamics is linked to enzyme catalysis

Katherine A. Henzler-Wildman¹, Ming Lei¹, Vu Thai¹, S. Jordan Kerns¹, Martin Karplus^{2,3} & Dorothee Kern¹

The synergy between structure and dynamics is essential to the function of biological macromolecules. Thermally driven dynamics on different timescales have been experimentally observed or simulated, and a direct link between micro- to milli-second domain motions and enzymatic function has been established^{1–4}. However, very little is understood about the connection of these functionally relevant, collective movements with local atomic fluctuations, which are much faster. Here we show that pico- to nano-second timescale atomic fluctuations in hinge regions of adenylate kinase facilitate the large-scale, slower lid motions that produce a catalytically competent state. The fast, local mobilities differ between a mesophilic and hyperthermophilic adenylate kinase, but are strikingly similar at temperatures at which enzymatic activity and free energy of folding are matched. The connection between different timescales and the corresponding amplitudes of motions in adenylate kinase and their linkage to catalytic function is likely to be a general characteristic of protein energy landscapes.

Enzymatic activity is known to require a precise balance between flexibility and stability. However, a key question remains: how do protein motions on different timescales relate to each other and contribute to this balance? To function, enzymes must be stable enough to retain their native three-dimensional structures, but flexible enough to allow sufficient substrate binding, chemical reaction and product release. Recent work on adenylate kinase (Adk) has shown that collective domain motion on the μ s–ms timescale—in particular, opening of the nucleotide binding lid(s) required for product release—is the rate-limiting step in catalysis for both a mesophilic (*Escherichia coli*, mesoAdk) and a hyperthermophilic (*Aquifex aeolicus*, thermoAdk) homologue, and that the difference in the lid-opening rate fully accounts for the difference in activity between them¹. Furthermore, lid motions along the opening–closing trajectory occur on similar timescales even in the absence of substrate in thermoAdk⁴. The linkage between dynamics and catalysis has been demonstrated primarily on the μ s–ms timescale, which correlates directly with the enzymatic turnover rate^{1,2}. However, thermal fluctuations of individual atoms occur much faster than this. This idea of a hierarchy of substates was originally proposed in work on myoglobin³. Mössbauer, infrared and neutron scattering measurements on myoglobin led to the interesting suggestion that some aspects of protein dynamics are ‘slaved’ to solvent fluctuations, with the protein component dictating the relative rates^{5,6}. In this work, we directly investigate the role of the protein component at atomic resolution and demonstrate a connection between ps–ns backbone fluctuations and the slower motions associated with catalytic activity.

To address this, we measured ¹⁵N amide bond dynamics by nuclear magnetic resonance (NMR) for meso- and thermoAdk as a function of temperature, and computed them with molecular dynamics

(MD). Previous comparisons of activity and dynamics in mesophilic/thermophilic homologues have been based primarily on H/D exchange experiments^{7,8}. These results generally show decreased overall exchange rates in the thermophilic protein at ambient temperatures relative to the mesophilic homologues, and for alcohol dehydrogenase indicated a possible uncoupling of activity and stability⁷. Region-specific variation in the dynamics between thermophilic and mesophilic RNase H have been observed in an NMR study of ps–ns backbone fluctuations⁹, whereas a uniform trend was observed by H/D exchange⁸. Here we characterize the hierarchy in dynamic timescales with respect to enzyme function.

NMR relaxation data (R_1 , R_2 , and ¹H–¹⁵N heteronuclear overhauser enhancement (NOE)) were analysed using the model-free approach^{10–12}, which allows extraction of the internal motional parameters (τ_c , internal correlation time; and S^2 , order parameter) for each backbone amide, plus the overall tumbling parameters. S^2 reflects the amplitude of amide bond fluctuation on the ps–ns timescale, ranging from 0 (freely jointed motion) to 1 (completely rigid). This analysis is complicated if the protein samples multiple conformations on a μ s–ms timescale, but our extensive knowledge of the substates and corresponding rates of interconversion between open and closed states in ligand-free Adk⁴ enabled us to dissect out the contributions of fast motions to the measured relaxation rates (see Methods).

Comparison of the homologues at 20 °C reveals that mesoAdk has significantly lower-order parameters than thermoAdk (Fig. 1a, b), consistent with the large difference in thermal stability (T_m is 54 °C for mesoAdk¹³ and 109 °C for thermoAdk (Supplementary Fig. 9a)). The average order parameter decreases with increasing temperature in thermoAdk (Fig. 2), as expected for increasing thermal atomic fluctuations¹⁴. Interestingly, similar order parameters are observed for thermo- and meso-Adk at temperatures equidistant below T_m (Figs 1a and 2d), supporting the hypothesis that these ps–ns backbone fluctuations are directly related to structural stability. To investigate this relationship quantitatively, we measured the free energy of unfolding (ΔG) by circular dichroism spectroscopy for mesoAdk at 20 °C and thermoAdk at 20 °C and 70 °C (Supplementary Fig. 9b–d). Indeed, the overall stability of mesoAdk at 20 °C ($\Delta G = 7.4 \pm 0.6$ kcal/mol; mean \pm s.d.) is identical to thermoAdk at 70 °C ($\Delta G = 7.4 \pm 0.6$ kcal/mol; mean \pm s.d.), in contrast to a ΔG of 23 ± 2 kcal mol^{–1} (mean \pm s.d.) for thermoAdk at 20 °C.

The observed rigidification in a number of regions in thermoAdk at 20 °C (Fig. 1a, b) suggests a ‘distributed’ mechanism of adaptation to extreme temperatures^{8,15}. This agrees with mutational studies (D. Kern *et al.*, unpublished data), and domain swapping experiments on another pair of Adk homologues¹⁶ in which the major contribution comes from the core region. This supports the idea that the thermostability of thermoAdk arises from the combined effect of multiple interactions. Because order parameters are most sensitive to

¹Department of Biochemistry, Howard Hughes Medical Institute, Brandeis University, Waltham, Massachusetts 02454, USA. ²Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ³Laboratoire de Chimie Biophysique, Institut de Science et d’Ingénierie Supramoléculaires, Université Louis Pasteur, 8 rue Gaspard Monge, F-67000 Strasbourg, France.

local packing¹⁷, we tested whether a coarse-grained method, FIRST (Floppy Inclusion and Rigid Substructure Topography)¹⁸, would sense the subtle factors that collectively account for the difference in stability. This method uses a network model to calculate degrees of freedom for each atom. The results (Fig. 1c, d) agree well with the experimental order parameters, demonstrating that small variations that alter the density of contacts are the main cause for the observed difference in order parameters and hence fast-timescale fluctuations, as seen for crystallographic B-factors¹⁹.

If increased ps–ns flexibility correlates with reduced stability, why have proteins evolved to be so dynamic? It is widely accepted that flexibility is crucial for activity and this has been demonstrated for μ s–ms motions², but the specific connection of these slow domain motions to fast, local fluctuations has not been established. Inspection of S^2 for both mesoAdk and thermoAdk reveals several flexible ‘hotspots’. Strikingly, many are located exactly where local backbone conformation must change for lid motion to occur (arrows in Figs 1–3). These hinges were computationally identified from differences in both pseudo-angles (between every three consecutive $C\alpha$ atoms) and pseudo-dihedral angles (between every four consecutive $C\alpha$ atoms) between the open and fully closed states (Fig. 4d, and Supplementary Fig. 10). Although most hinges are similar to those previously identified in mesoAdk²⁰, the novel calculation of pseudo-angle changes identified a new hinge (5) with scissor-type motion (Supplementary Table 3 and Supplementary Fig. 10).

Our NMR results suggest that local, ps–ns fluctuations are involved in facilitating the slower, larger-amplitude domain motions

as predicted^{4,21,26}. Complex motional modes on multiple timescales greatly complicate model-free analysis and lead to either complex model selection or even failure of model-free. This is indeed detected for residues in the lids and hinges of thermoAdk (Fig. 4a), substantiating this hypothesis. Specifically, fast motions on two timescales (ps and about 1 ns) are detected in addition to global tumbling and the micro- to milli-second lid opening/closing. Interestingly, the same pattern is observed in mesoAdk (Fig. 4b) even though multiple conformations were not captured in the crystal structure. A recent principal component analysis on mesoAdk²² also found lid motions.

MD simulations provide another method for assessing ps–ns dynamics at atomic resolution. They provide a complete set of order parameters, overcoming the problem of incomplete experimental data, they address questions of motional coupling^{23,24} and they enable cross-validation with experiments. Amide order parameters calculated from the MD trajectories⁴ agree very well with the experimental data (Fig. 3). Increased amide fluctuations at higher temperature are seen for all hinges (Fig. 3b–d), completing the experimentally observed trend that was limited by the loss of amide signals owing to exchange with water.

One interesting hinge (8) is located where the long helix kinks on lid closure. Normally, small-amplitude amide bond fluctuations are observed in stable helices; slightly decreased order parameters (Fig. 1) at this site provide evidence for a weak spot. If the distinctive properties of this site are important for function it should be evolutionarily conserved. An alignment of 250 Adk sequences reveals strong conservation of this and most other hinges, comparable to the conservation of the active site (Fig. 4c–e, and Supplementary Fig. 11). Moreover, a cluster of residues with low helical propensity around hinge 8 is also conserved, including a proline residue, possibly rationalizing the reduced order parameters. Local unfolding, referred to as cracking, has been proposed to lower the barrier of opening/closing in mesoAdk²⁵. However, our experimental results identify different spots

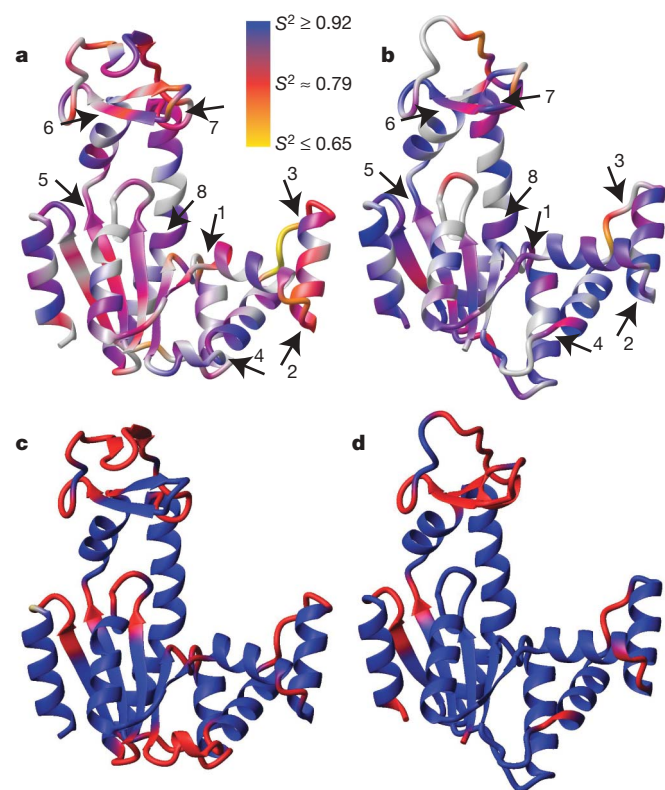


Figure 1 | Comparison of thermoAdk and mesoAdk flexibility. **a, b,** S^2 values are shown for mesoAdk (**a**) and thermoAdk (**b**) at 20 °C as a continuous colour scale with grey for both prolines and residues for which S^2 cannot be measured owing to spectral overlap or fast exchange of the amide protons with water. Hinges are numbered and highlighted with arrows. **c, d,** Computational analysis using FIRST¹⁸ captures the experimental patterns of flexibility in mesoAdk (**c**) and thermoAdk (**d**); blue, rigid regions with more contacts than degrees of freedom; red, residues have more degrees of freedom than contacts; magenta, residues have the same number of restraining contacts and degrees of freedom.

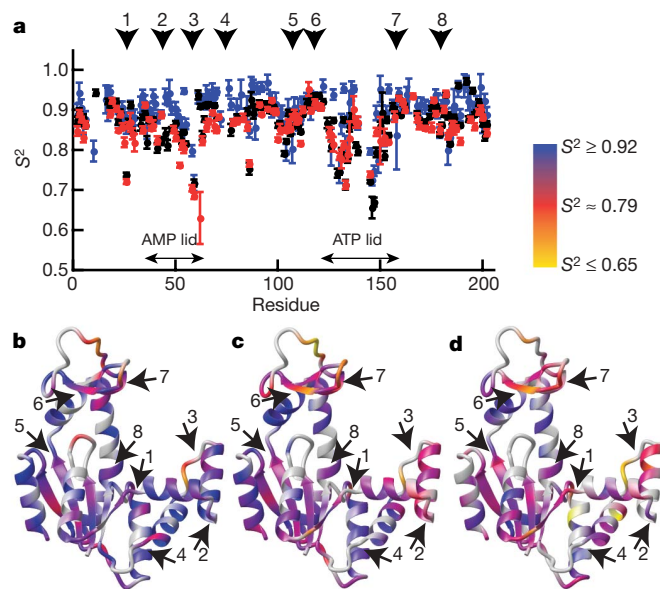


Figure 2 | Temperature dependence of thermoAdk order parameters. **a,** S^2 values at 20 °C (blue), 50 °C (black) and 80 °C (red; 80 °C sample contains 20% glycerol to reduce overall tumbling time, see Supplementary Fig. 8). **b, c, d,** These data were plotted onto the thermoAdk crystal structure at 20 °C (**b**), 50 °C (**c**) and 80 °C (**d**) using the same colour scale as in Fig. 1, and arrows to indicate hinges. The decrease of S^2 with temperature reflects increased flexibility on the ps–ns timescale, with similar S^2 values for thermoAdk at 80 °C (**d**) as mesoAdk at 20 °C (Fig. 1a). Because nearly identical order parameters were obtained in the model-free analysis using each of the three apo thermoAdk conformations present in the crystal structure⁴ (Supplementary Fig. 5), all figures show the results using only molecule A (the most open conformation) unless otherwise noted. Error bars, 1 s.d.

of increased ps–ns motions to those suggested in ref. 25, and agree with the hinge regions determined on the basis of the structures of the open and closed state. The reduced order parameters suggest that this hinge is already flexible in the fully open apo state.

Interestingly, most hinges have higher order parameters in thermoAdk than in mesoAdk at 20 °C (Fig. 1a, b). Because these hinges must rearrange for lid opening, which is the source of slower turnover rates for thermoAdk, one can suggest that the smaller amplitudes of the local, fast-timescale motions could be the origin of the measured differences in the large-scale microsecond motions. This hypothesis is supported by similar order parameters of the hinges in thermoAdk and mesoAdk at temperatures of comparable activities (Figs 1a, 2d and 3b, d). Examination of sequence and structure in the hinge regions suggests possible reasons for differences in backbone flexibility. Against the highly conserved background of the hinge regions, there are single non-conservative amino acid substitutions within a hinge or spatially close by in thermoAdk relative to mesoAdk (Fig. 4f, and Supplementary Fig. 10c). It is precisely this fact that indicates that these changes might have an important role in thermal adaptation. First, four of the hinges (2, 3, 4 and 7) contain proline only in thermoAdk, a residue that tends to rigidify the backbone. Second, aromatic ring-stacking interactions are found near hinges 4 and 5 in thermoAdk, but not in mesoAdk (Fig. 4f).

To test our hypothesis that these amino acid differences are the source of increased picosecond local flexibility of mesoAdk in the hinges, we made use of the good agreement between experimental and computational order parameters by simulating a sextuple mutant of thermoAdk. In this mutant, four prolines and two phenylalanines in the hinge regions of thermoAdk were replaced with the corresponding amino acids of mesoAdk (P44E, P60T, P73A, P155E, F5L and F80L). Relative to wild-type thermoAdk, the sextuple mutant indeed shows lower order parameters at hinges 2, 4 and 7 (Supplementary Fig. 13). The comparison between these computed differences in order parameters on mutation with the experimentally measured differences between mesoAdk and thermoAdk (Supplementary Fig. 13) suggest that these point mutations in the otherwise highly conserved hinges are the physical origin for the observed differences in motional

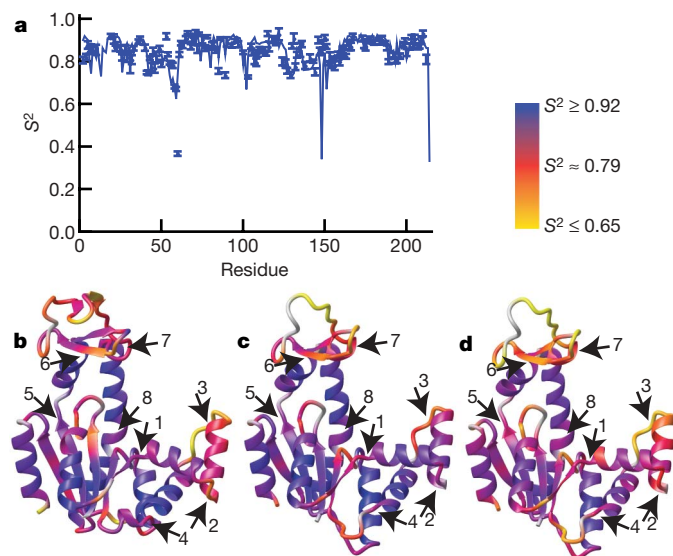


Figure 3 | Temperature dependence of MD order parameters and comparison with experimental S^2 . **a**, Order parameters were calculated from 10-ns MD simulations for mesoAdk at 300 K (**a**, blue line) and thermoAdk at MD temperatures of 300 K and 350 K; the MD results agree well with the experimental data for mesoAdk at 293 K (**a**, blue circles) and thermoAdk (Supplementary Fig. 13a). **b**, **c**, **d**, The S^2 values calculated from MD simulations are plotted onto the structures for mesoAdk at 300 K (**b**) and thermoAdk at 300 K (**c**) and 350 K (**d**). Error bars, 1 s.d.

amplitudes. Interestingly, although these mutations affect the amplitude of motion, they do not seem to alter the directionality of motion (Fig. 4g–i), which is along the opening–closing trajectory. Thus the functionally important directionality is preserved while the amplitude is modified by sequence and temperature.

The comparative analysis of the dynamics between a thermophilic and mesophilic enzyme at atomic resolution enabled identification of atomic fluctuations that are crucial for activity. NMR relaxation techniques provided the unique opportunity to measure picosecond

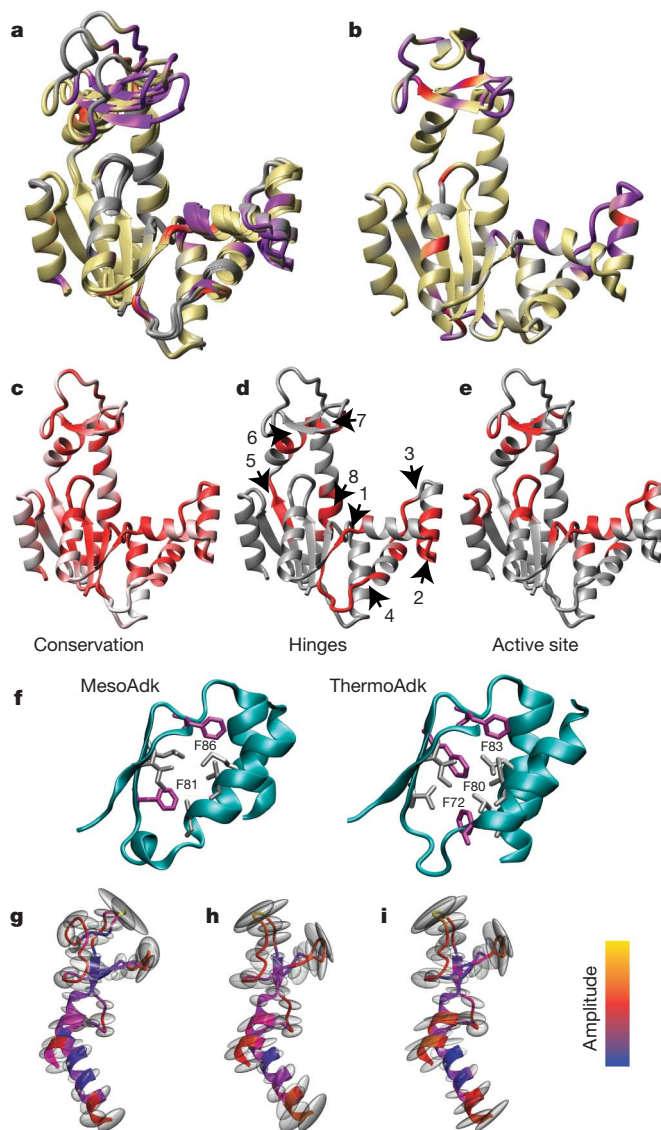


Figure 4 | Linkage of motional timescales and characterization of hinges. **a**, **b**, Residues that require multiple fast-timescale dynamics (model 5 or 6, see Methods) to fit the NMR relaxation data (purple) or cannot be fit using model-free analysis (red) are shown for thermoAdk at 50 °C (**a**) and mesoAdk at 20 °C (**b**). Complex dynamics in these regions result from lid fluctuations as suggested by the three lid conformations captured in the thermoAdk crystal structure (overlayed in **a**) and MD calculations⁴. **c**, The conservation score based on a ClustalW²⁸ alignment of 250 Adk sequences (Supplementary Fig. 11) is mapped onto the structure of thermoAdk, coloured from white to red with increasing conservation. **d**, **e**, The hinges, numbered and coloured in red on the structure (**d**), are as highly conserved as active site residues (**e**, residues within 5 Å of substrate in the closed conformation are coloured red). **f**, Packing differences near hinge 4 in mesoAdk and thermoAdk caused by different aromatic ring stacking interactions. The direction of motion along the opening–closing trajectory is preserved, as shown by the directionality of fluctuations of the 30 lowest frequency principal components³⁰ around hinge 6 and 7 in mesoAdk (**g**), thermoAdk (**h**), and the sextuple mutant of thermoAdk (**i**; in this region the only mutation is P155E).

to millisecond motions throughout the protein, thus experimentally connecting the various tiers of the energy landscape⁴. The correspondence of ps–ns flexibility in the hinges between the mesophilic and hyperthermophilic Adk at temperatures at which enzymatic activities are matched provides the link between local, fast-timescale dynamics and slower, global lid-opening/-closing. Moreover, the combination of structural and dynamics data point towards specific atomic origins for the differences in activity/dynamics between these two homologous enzymes, the rates of which are limited by the lid motions. This hypothesis was substantiated by molecular dynamics simulations on *in silico* mutants. We conjecture that the ‘slowness’ of lid opening/closing is in part due to the requirement of collective motions of many atoms within each hinge and the correlated motion of several hinges. In other words, closing is an improbable, hence slow, event, arising from many individual attempts by local groups to overcome the energy barrier for the conformational transition. Of course all atoms in these regions, backbone and side chain, participate collectively; here we have only measured atomic dynamics for the backbone amides as representative markers because the conformational changes linked to function involve significant backbone rearrangement⁴. The role of the hinges in linking local fluctuations with functionally critical conformational transitions distinguishes regions that contribute primarily to activity, and the high ps–ns flexibility at these sites permits the hinge re-orientation to occur with a lower barrier²⁶. The nature of interatomic interactions developed by evolution in proteins underpins their unique ability to sample specific conformations efficiently within the folded subspace.

METHODS SUMMARY

Sample preparation. ¹⁵N-labelled thermo- and meso-Adk were expressed and purified as described¹. NMR samples were 1 mM Adk, 40 mM MOPS, 50 mM NaCl, 10% D₂O, pH 7.0. Circular dichroism samples were prepared in the same buffer with protein concentrations of 10–25 μM.

NMR experiments and model-free analysis. *R*₁, *R*₂ and heteronuclear NOE data were acquired using standard pulse sequences on Varian Innova 500 and 600 MHz spectrometers. Multiple NOE experiments were averaged for each sample. Fast-ModelFree¹⁰ and ModelFree 4.15¹² were used to optimize the fit of both internal dynamics and global tumbling parameters, using a data set filtered to remove residues with low signal-to-noise ratios and those with *R*_{ex} contributions to *R*₂, as measured with CPMG experiments⁴. Dynamic parameters for all residues were then fit using the optimized axially symmetric diffusion tensor and the known *R*_{ex} values⁴.

Computation. FIRST was performed as described¹⁸ with the default hydrogen bond energy cut off value of −1.0 kcal mol^{−1}. FIRST predicts protein flexibility by counting the local density of covalent bonds, hydrogen bonds and hydrophobic interactions. MD simulations were performed as described⁴. Briefly, molecule A of thermoAdk and 4AKE (Protein Data Bank ID code) for mesoAdk were simulated in explicit solvent for 10 ns, with a snapshot saved every picosecond. Every snapshot was superimposed onto a reference frame to remove the contribution of overall molecular motion. Then order parameters were calculated from the MD trajectory using the NMR module in CHARMM²⁷. Hinges were calculated on the basis of a combination of changes in pseudo-angle and pseudo-dihedral-angle between the fully open and fully closed conformations of mesoAdk or thermoAdk (Supplementary Fig. 10 and Supplementary Table 3). To assess sequence homology, 250 Adk sequences closely related to thermoAdk, with archaea Adks excluded, were aligned using CLUSTALW²⁸.

All structure figures were prepared using molmol²⁹.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 April; accepted 29 October 2007.

Published online 18 November 2007.

1. Wolf-Watz, M. *et al.* Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nature Struct. Mol. Biol.* **11**, 945–949 (2004).
2. Boehr, D. D., Dyson, H. J. & Wright, P. E. An NMR perspective on enzyme dynamics. *Chem. Rev.* **106**, 3055–3079 (2006).
3. Austin, R. H., Beeson, K. W., Eisenstein, L., Frauenfelder, H. & Gunsalus, I. C. Dynamics of ligand binding to myoglobin. *Biochemistry* **14**, 5355–5373 (1975).
4. Henzler-Wildman, K. *et al.* Intrinsic motions along an enzymatic reaction trajectory. *Nature* advanced online publication doi:10.1030/nature06410 (18 November 2007).

5. Vitkup, D., Ringe, D., Petsko, G. A. & Karplus, M. Solvent mobility and the protein ‘glass’ transition. *Nature Struct. Biol.* **7**, 34–38 (2000).
6. Fenimore, P. W., Frauenfelder, H., McMahon, B. H. & Parak, F. G. Slaving: Solvent fluctuations dominate protein dynamics and functions. *Proc. Natl Acad. Sci. USA* **99**, 16047–16051 (2002).
7. Nagel, Z. D. & Klinman, J. P. Tunneling and dynamics in enzymatic hydride transfer. *Chem. Rev.* **106**, 3095–3118 (2006).
8. Hollen, J. & Marqusee, S. Structural distribution of stability in a thermophilic enzyme. *Proc. Natl Acad. Sci. USA* **96**, 13674–13678 (1999).
9. Butterwick, J. A. *et al.* Multiple time scale backbone dynamics of homologous thermophilic and mesophilic ribonuclease HI enzymes. *J. Mol. Biol.* **339**, 855–871 (2004).
10. Cole, R. & Loria, J. P. FAST-ModelFree: A program for rapid automated analysis of solution NMR spin-relaxation data. *J. Biomol. NMR* **26**, 203–213 (2003).
11. Lipari, G. & Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546–4559 (1982).
12. Mandel, A. M., Akke, M. & Palmer, A. G. Backbone dynamics of *Escherichia coli* ribonuclease H—correlations with structure and function in an active enzyme. *J. Mol. Biol.* **246**, 144–163 (1995).
13. Monnot, M. *et al.* Circular-dichroism investigation of *Escherichia coli* adenylate kinase. *J. Biol. Chem.* **262**, 2502–2506 (1987).
14. Mandel, A. M., Akke, M. & Palmer, A. G. Dynamics of ribonuclease H: Temperature dependence of motions on multiple time scales. *Biochemistry* **35**, 16009–16023 (1996).
15. Petsko, G. A. Structural basis of thermostability in hyperthermophilic proteins, or ‘there’s more than one way to skin a cat’. *Methods Enzymol.* **334**, 469–478 (2001).
16. Bae, E. & Phillips, G. N. Roles of static and dynamic domains in stability and catalysis of adenylate kinase. *Proc. Natl Acad. Sci. USA* **103**, 2132–2137 (2006).
17. Zhang, F. L. & Bruschweiler, R. Contact model for the prediction of NMR ¹⁵N-¹H order parameters in globular proteins. *J. Am. Chem. Soc.* **124**, 12654–12655 (2002).
18. Jacobs, D. J., Rader, A. J., Kuhn, L. A. & Thorpe, M. F. Protein flexibility predictions using graph theory. *Proteins* **44**, 150–165 (2001).
19. Halle, B. Flexibility and packing in proteins. *Proc. Natl Acad. Sci. USA* **99**, 1274–1279 (2002).
20. Maragakis, P. & Karplus, M. Large amplitude conformational change in proteins explored with a plastic network model: Adenylate kinase. *J. Mol. Biol.* **352**, 807–822 (2005).
21. Clore, G. M. & Schweiters, C. D. Concordance of residual dipolar coupling, backbone order parameters and crystallographic B-factors for a small α/β protein: A unified picture of high probability, fast atomic motions in proteins. *J. Mol. Biol.* **355**, 879–886 (2006).
22. Lou, H. F. & Cukier, R. I. Molecular dynamics of apo-adenylate kinase: A principal component analysis. *J. Phys. Chem. B* **110**, 12796–12808 (2006).
23. Prompers, J. J. & Bruschweiler, R. Thermodynamic interpretation of NMR relaxation parameters in proteins in the presence of motional correlations. *J. Phys. Chem. B* **104**, 11416–11424 (2000).
24. Temiz, N. A., Meirovitch, E. & Bahar, I. *Escherichia coli* adenylate kinase dynamics: Comparison of elastic network model modes with mode-coupling ¹⁵N-NMR relaxation data. *Proteins* **57**, 468–480 (2004).
25. Miyashita, O., Onuchic, J. N. & Wolynes, P. G. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc. Natl Acad. Sci. USA* **100**, 12570–12575 (2003).
26. Karplus, M. & McCammon, J. A. The internal dynamics of globular-proteins. *CRC Crit. Rev. Biochem.* **9**, 293–349 (1981).
27. Brooks, B. R. *et al.* CHARMM - a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4**, 187–217 (1983).
28. Chenna, R. *et al.* Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **31**, 3497–3500 (2003).
29. Koradi, R., Billeter, M. & Wuthrich, K. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55 29–32 (1996).
30. Karplus, M. & Kushick, J. N. Method for Estimating the Configurational Entropy of Macromolecules. *Macromolecules* **14**, 325–332 (1981).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Kay for providing pulse programs. We are grateful to K. O. Stetter for providing DNA isolated from *Aquifex aeolicus* and the Advanced Biomedical Computing Center for CPU hours. This work was supported by NIH grants (D.K. and K.A.H.-W.), a DOE grant (D.K.) and a fellowship from the American Heart Association (M.L.). The research at Harvard was supported in part by a grant from NIH to M.K.

Author Contributions K.A.H.-W. and M.L. contributed equally to this work. K.A.H.-W. performed the NMR experiments with supervision from D.K., and M.L. carried out the computational analysis with supervision from M.K. and D.K. All authors discussed the results and interpretation. D.K. and K.A.H.-W. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.K. (dkern@brandeis.edu).

METHODS

NMR model-free analysis. The NMR data were fitted using model-free analysis^{10,12} with statistical selection among the 5 standard models, plus an additional model to account for selection of model 5 in combination with a known R_{ex} value. For each model, the parameters required to fit the data are: model 1, S^2 ; model 2, S^2 , $\tau_e = \tau_f$ (τ_f is fast, low ps); model 3, S^2 , R_{ex} ; model 4, S^2 , $\tau_e = \tau_f$, R_{ex} ; model 5, S^2 , S^2 , $\tau_e = \tau_s$ (τ_s is slow, hundreds of picoseconds to low nanosecond values); and model 6, S^2 , S^2 , $\tau_e = \tau_s$, R_{ex} .

The impact of complex motions. Analysis of ps–ns dynamics from NMR data in the presence of slower domain motions is complicated by several factors: R_2 data are sensitive to motions on the μ s–ms timescale, large-scale conformational changes in conjunction with axially symmetric anisotropic overall diffusion lead to re-orientation of the N–H bond with respect to the diffusion tensor, and coupled motions may violate assumptions implicit in model-free analysis. We used several approaches to determine the significance of these factors and corrected for them when necessary.

First, we removed μ s–ms-timescale exchange contributions (R_{ex}) that are due to lid movement from the observed R_2 values, using the exact R_{ex} values independently determined by CPMG experiments⁴ before model-free analysis. R_1 and NOE data were measured at 500 and 600 MHz, but R_2 data were only obtained for a single field (500 MHz) because R_{ex} increases and becomes more difficult to correct for at higher fields, and greater inconsistencies have been observed in R_2 data³¹.

Second, we confirmed that reorientation of the N–H bond vectors with respect to the diffusion tensor that is due to lid movement does not significantly affect the extraction of the order parameters. Each of the three conformations found in the crystal structure of apo thermoAdk were used for fitting the diffusion tensor during model-free analysis of the NMR relaxation data collected for apo thermoAdk. The optimized diffusion tensors are very similar for each of the conformations analysed, so the orientation of the amides in the core does not change significantly with respect to the diffusion tensor, whereas the amides in the lids have different relative orientations. Calculation of the order parameters with each of the three substates results in order parameters that are the same within error (Supplementary Fig. 5 shows the results at 20 °C).

Third, order parameters calculated from MD simulations (using the CMAP correction to CHARMM22 that is crucial for correct determination of RMSF fluctuations in MD³²), which sampled a wide range of lid conformations, showed striking agreement with the experimental values (Fig. 3). This demonstrates that possible mode coupling between global, local and domain motions^{24,33,34} or correlated motions did not distort the derivation of the experimentally determined local fluctuations (S^2) from model-free, which neglects motional coupling. The conformational space covered during the 10-ns MD trajectories is larger than that captured by the three molecules in the crystal structure. As a result, S^2 values calculated from the MD trajectories include the effects of the correlated motion of the lids and are independent of global tumbling.

Final confirmation that the internal dynamic parameters were well optimized was established by comparing fitted R_{ex} values from the model-free output with those directly measured from experiments. Non-isotropic diffusion and the effects of μ s–ms dynamics can be difficult to distinguish, with extraneous R_{ex} terms appearing in model-free output if an incorrect diffusion tensor is used³⁵. With the final optimized diffusion tensor, R_{ex} terms were only required to fit residues with known μ s–ms motion or in loop regions for which microsecond dynamics are likely. R_{ex} is not well determined by model-free analysis and the values did not match exactly, but correction of initial R_2 using CPMG experiments, as described in the main text, eliminated the need for almost all R_{ex} terms when fitting the data.

These results suggest that assumptions made during model-free analysis of the experimental data are valid: global diffusion can be modelled as an axially symmetric diffusion tensor using a single average structure for apo thermoAdk and the correlated motion of the lid regions about the average open conformation does not significantly affect the extraction of order parameters from NMR relaxation data using model-free, as described. The average conformation during the MD trajectories was very close to the structure of molecule A in the crystal⁴, and therefore this conformation was used for final analysis of the data presented here. **Minor population of the closed state of apo Adk.** The MD trajectories and NMR relaxation parameters both probe the same ps–ns timescale. Domain motion on a much slower timescale may interfere with interpretation of the NMR relaxation data, particularly if states with very different rotational diffusion properties are significantly populated. NMR CPMG relaxation dispersion and single molecule FRET experiments⁴ suggest that apo thermoAdk does sample a fully closed state to a small percentage but the exact population could not be determined. The consistency of the order parameters determined by MD and NMR indicates that the relative population of the closed state must be low, such that it does not disturb extraction of the local order parameters.

To address this question independently, the global correlation time of thermoAdk was re-examined to investigate whether the presence of fully closed conformations was likely to skew analysis of the internal dynamics. Optimization of τ_c during model-free analysis of apo thermoAdk at 20 °C produced a final fitted value of 18.2 ns. Model-free analysis of fully closed thermoAdk (saturated with Mg•Ap5A) resulted in an optimized τ_c of 14.0 ns at 20 °C (Supplementary Fig. 7). Calculation of τ_c based only on the crystal structure (using molecule A for open and Mg•Ap5A bound for closed) was performed using HYDRONMR^{36,37}, which uses a bead model to determine the hydrodynamic properties for a given solvent viscosity (temperature is indirectly specified through the water viscosity). The atomic element radius used for the bead model varies slightly from protein to protein depending on the exact hydration properties of the surface, with a narrow range from 2.8–3.8 for most proteins. For both apo and Mg•Ap5A-bound thermoAdk, this was found to be 3.2, well within the normal range. If the structure used for HYDRONMR calculations is not a good model of the solution state, significant deviations in atomic element radius would be expected. Strikingly, the atomic element radius of 3.2 resulted in calculated τ_c values of 18.2 ns for apo thermoAdk and 14.0 ns for Mg•Ap5A-bound thermoAdk, exactly matching the experimental τ_c values. Significant population of the closed state in apo thermoAdk should lead to a mismatch, with the experimental τ_c lower than the calculated value. Using the NMR relaxation data from apo thermoAdk to represent the fully open conformation and Mg•Ap5A-bound thermoAdk to represent the fully closed state, the expected relaxation parameters for a 20% closed/80% open state were calculated, leading to a one nanosecond decrease in τ_c . This confirms that a 20% population of the closed state would cause an observable discrepancy between the calculated and experimental results and the actual fully closed population is smaller.

NMR relaxation parameters calculated with HYDRONMR reflect global diffusion, including anisotropic tumbling effects that are due to the structure, but ignore internal dynamics, assuming a completely rigid structure. Therefore, the experimental values should be less than the calculated values as a result of internal ps–ns fluctuations ($S^2 < 1$). This is true for R_1 values for all but a few loop residues (Supplementary Fig. 6a) in which the N–H vector orientation in the crystal is most likely to be a poor model of the solution conformation. Of the six residues with $R_2^{NMR} > R_2^{CALC}$ (Supplementary Fig. 6b), three have detectable R_{ex} contributions, based on relaxation dispersion experiments, and the remaining three are again in loops. These results, particularly the close correspondence between τ_c calculated from the structure and the NMR data, strongly support only very minor population of the fully closed state and indicate that molecule A is a good model of the average solution conformation of apo thermoAdk for extraction of individual order parameters. The anisotropy of the axially symmetric diffusion tensor ($D_{||}/D_{\perp}$) is nearly constant for all apo thermoAdk samples (Supplementary Table 1) and τ_c fits the Stokes–Einstein equation over the entire temperature range, with a single hydrodynamic radius of 26.1 Å, indicating that there is not a significant change in the average degree of lid closure or relative populations of the various conformations present in solution over this temperature range (Supplementary Fig. 7). For the Mg•Ap5A-bound closed state, a smaller τ_c and a constant hydrodynamic radius (23.8 Å) with temperature was measured (Supplementary Fig. 7). This internal consistency of the global tumbling parameters across the wide temperature range accessible with thermoAdk also provides an internal check on the accuracy of the optimized diffusion model.

The problem of separability of dynamic time scales. Initial model-free analysis of thermoAdk at 80 °C found that the order parameters in the core decreased continually with temperature, as expected, but the average order parameter in the AMP-lid remained the same between 50 °C (Fig. 2c) and 80 °C (Supplementary Fig. 8b), and two of the helices even seemed to become more rigid at 80 °C. Previous studies^{14,38–43} have found a small but observable decrease in order parameter with increasing temperature in secondary structure elements, with larger decreases in S^2 in loops or at temperatures approaching the denaturation temperature. Closer inspection of the data suggested that this discrepancy in the lid regions probably resulted from breakdown in the assumption of the decoupling approximation at high temperatures. Model-free analysis collapses when the decoupling approximation, $\tau_e/\tau_c \ll 1$, is violated⁴⁴. This is only a problem for residues that require model 5, which includes a slower internal correlation time (τ_e , hundreds to thousands of picoseconds) that may approach the global correlation time (τ_c , nanoseconds), in addition to a fast internal correlation time (less than tens of picoseconds). These residues are found primarily in the lids as discussed in the main text (Fig. 4a, b). At 50 °C, model 5 residues in the AMP lid have $\tau_e/\tau_c \approx 0.1$ (Supplementary Fig. 8a). Violation of the decoupling approximation leads to overestimation of both S^2 and τ_e , so at 50 °C the order parameters are likely to be slightly overestimated (≤ 5 –10%)⁴⁴. Correlation times for internal motion (τ_e) have not been found to decrease greatly with temperature^{14,45}, but the overall tumbling of the protein (τ_c , ns) does become much

faster as the viscosity of water decreases. Under these conditions, separability of global and internal motions will be reduced at higher temperatures, leading to incorrect model selection and overestimation of order parameters^{14,44,45}. This is exactly what is observed at 80 °C; multiple fast-timescale internal dynamics can no longer be detected, leading to incorrect selection of model 1 (Supplementary Table 2) and significant overestimation of the order parameters (Supplementary Fig. 8b).

To accurately analyse the internal dynamics, the NMR relaxation data were re-measured at 80 °C with 20% glycerol added to the sample (Supplementary Fig. 8a, c) to reduce the global correlation time. An analogous strategy was recently employed to study RNA dynamics by elongating the RNA to slow and decouple global rotation⁴⁶. The core of the protein, which has simpler internal dynamics, serves as a control, confirming that glycerol does not affect the fast-timescale internal dynamics (Supplementary Fig. 8b, c). This strategy permits the multiple ps–ns internal motions of the lids to be accurately characterized at the highest temperature, demonstrating that the trend of decreasing order parameter with increasing temperature is maintained throughout thermoAdk, including the lid and hinge regions, as discussed above.

Chemical denaturation by circular dichroism spectroscopy. MesoAdk or thermoAdk samples were incubated overnight at the specified temperature and guanidine hydrochloride concentrations. Circular dichroism spectra were measured on a Jasco-710 spectropolarimeter equipped with a peltier cell holder and a PTC-348 temperature controller. An average of five scans from 230–215 nm were taken for each sample, and each data point is an average of three samples. The mean residual molar ellipticity at 222 nm was fitted to a two-state model with linear extrapolation. Dilution of unfolded Adk (6 M guanidine hydrochloride) into excess buffer was used to confirm the reversibility of denaturation.

Stability of thermoAdk by differential scanning calorimetry. To determine the T_m of thermoAdk (Supplementary Fig. 9a), differential scanning calorimetry (DSC) was performed with a VP-DSC microcalorimeter (MicroCal). ThermoAdk at 200 $\mu\text{g ml}^{-1}$ in 50 mM sodium cacodylate, 80 mM KCl, pH 7.4, was heated from 25 °C to 120 °C at a rate of 90 °C h⁻¹ under a constant pressure of 1.8 atm. Matching buffer was used as a reference. Thermal denaturation of thermoAdk was only partly reversible, as assessed by repeated scans. Buffer–buffer scans were performed and subtracted from the protein–buffer scans to minimize noise.

Computation. The NMR module in CHARMM²⁷ calculates the auto-correlation function of the second Legendre polynomial of every N–H bond vector, and the order parameter is taken as the long-time limit of this autocorrelation function¹¹. The CMAP backbone energy correction in the CHARMM force field was included; it has been shown to significantly improve the accuracy of order parameter computations³². The sextuple mutant thermoAdk was constructed in CHARMM and simulated by the same protocol as the wild-type thermoAdk. To address the question of directionality of motions around the hinges, a principal component analysis was performed for hinge regions by diagonalizing the equal time mass weighted atom fluctuation cross-correlation matrix³⁰. Only the fragment shown (Fig. 4g–i) was aligned to remove global translation and rotation when computing the principal components from the simulation trajectories. The first 30 modes were used to construct Fig. 4g–i because they have the largest contribution to the closing transition (Supplementary Fig. 13d).

Two methods were used to quantitatively locate the hinge regions using the open and closed crystal structures of mesoAdk (Protein Data Bank files: apo, 1AKE; Ap5A-bound, 4AKE) and thermoAdk (molecule A of the apo structure and the Ap5A-bound structure⁴). First, pseudo-angles between every three C α atoms and pseudo-dihedral-angles between every four C α atoms were calculated for each structure. Hinges were located wherever the difference between the open and closed form was greater than a cutoff value (Supplementary Fig. 10). For pseudo-angle (pseudo-dihedral-angle) changes the central one (two) residue(s) was assigned to the hinge. Second, dynamic domains and hinge regions were calculated using the program DynDom⁴⁷, which identifies domains on the basis of clustering of the rotation vectors used to transform segments of the protein chain from the open to the closed conformation. This method requires definition of a minimum domain size for determining how big a cluster is necessary to form a domain. Use of the default domain size of 20 residues results in three domains separated by four hinges, with each nucleotide lid and the core as a separate domain. Because several of the hinges identified above are much closer together, the DynDom analysis was repeated with a smaller minimum domain size. Identical results to the pseudo-angle/pseudo-dihedral-angle analysis were obtained with minimum domain limits between 7 and 10 residues, identifying all eight hinges in thermoAdk (Supplementary Table 3). In mesoAdk, seven hinges were identified, with the discrepancy at hinge 5/6. These hinges are too close together to be accurately separated with this method. Also, interaction of the

ATP lid with the ligand in the closed form of mesoAdk causes an additional kink in the helix between hinges 5 and 6, hinges that are identified in both thermoAdk and mesoAdk with the pseudo-angle analysis. Previous work has identified subsets of most of these hinges (Supplementary Table 3). The high homology of the hinge regions identified in mesoAdk and thermoAdk by multiple methods enabled us to construct a consensus definition of the hinge regions, which is shown on the sequence alignment (Fig. 3, and Supplementary Fig. 11).

Correlation between experimental and computed order parameters. The correlation plot demonstrates overall good agreement between experimental and computed order parameters (Supplementary Fig. 12a). For a small number of residues, the experimental S^2 are higher than the computed S^2 . These amides are located in loops, and close inspection of these outliers revealed that an over-simplified model has probably been selected by model-free analysis in these cases because neighbouring residues have more complex models resulting in lower order parameters, in agreement with the computed S^2 . On the other hand, there are a few residues for which the computed S^2 are higher than the experimental values. These amides are located in regions with significant nanosecond as well as picosecond fluctuations, as determined by NMR (see Fig. 4a, b). This over-estimation of computed S^2 is clearly a sampling problem owing to the limited length of the MD trajectories (10–ns trajectories). To see this more clearly, order parameters were calculated by superimposing five adjacent residues to eliminate nanosecond collective motional contributions to the relaxation calculated from the MD trajectories. Indeed, using this local rather than a global alignment leads to an overestimation of order parameters calculated from MD relative to the experimental values (see Supplementary Fig. 12b), which include both the local picosecond and the more collective nanosecond fluctuations.

- Lee, A. L. & Wand, A. J. Assessing potential bias in the determination of rotational correlation times of proteins by NMR relaxation. *J. Biomol. NMR* **13**, 101–112 (1999).
- Buck, M., Bouguet-Bonnet, S., Pastor, R. W. & MacKerell, A. D. Importance of the CMAP correction to the CHARMM22 protein force field: Dynamics of hen lysozyme. *Biophys. J.* **90**, L36–L38 (2006).
- Shapiro, Y. E. *et al.* Domain flexibility in ligand-free and inhibitor-bound *Escherichia coli* adenylate kinase based on a mode-coupling analysis of N-15 spin relaxation. *Biochemistry* **41**, 6271–6281 (2002).
- Tugarinov, V., Shapiro, Y. E., Liang, Z. C., Freed, J. H. & Meirovitch, E. A novel view of domain flexibility in *E. coli* adenylate kinase based on structural mode-coupling ¹⁵N NMR relaxation. *J. Mol. Biol.* **315**, 155–170 (2002).
- Hall, J. B. & Fushman, D. Characterization of the overall and local dynamics of a protein with intermediate rotational anisotropy: Differentiating between conformational exchange and anisotropic diffusion in the B3 domain of protein G. *J. Biomol. NMR* **27**, 261–275 (2003).
- Bernado, P., de la Torre, J. G. & Pons, M. Interpretation of ¹⁵N NMR relaxation data of globular proteins using hydrodynamic calculations with HYDRONMR. *J. Biomol. NMR* **23**, 139–150 (2002).
- de la Torre, J. G., Huertas, M. L. & Carrasco, B. HYDRONMR: Prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.* **147**, 138–146 (2000).
- Evenas, J., Forsen, S., Malmendal, A. & Akke, M. Backbone dynamics and energetics of a calmodulin domain mutant exchanging between closed and open conformations. *J. Mol. Biol.* **289**, 603–617 (1999).
- Seewald, M. J. *et al.* The role of backbone conformational heat capacity in protein stability: Temperature dependent dynamics of the B1 domain of Streptococcal protein G. *Protein Sci.* **9**, 1177–1193 (2000).
- Spyracopoulos, L. *et al.* Temperature dependence of dynamics and thermodynamics of the regulatory domain of human cardiac troponin C. *Biochemistry* **40**, 12541–12551 (2001).
- Zhuravleva, A. V. *et al.* Gated electron transfers and electron pathways in azurin: A NMR dynamic study at multiple fields and temperatures. *J. Mol. Biol.* **342**, 1599–1611 (2004).
- Chang, S. L. & Tjandra, N. Temperature dependence of protein backbone motion from carbonyl ¹³C and amide ¹⁵N NMR relaxation. *J. Magn. Reson.* **174**, 43–53 (2005).
- Wang, T. Z., Cai, S. & Zuiderweg, E. R. P. Temperature dependence of anisotropic protein backbone dynamics. *J. Am. Chem. Soc.* **125**, 8639–8643 (2003).
- Vugmeyster, L., Raleigh, D. P., Palmer, A. G. & Vugmeister, B. E. Beyond the decoupling approximation in the model free approach for the interpretation of NMR relaxation of macromolecules in solution. *J. Am. Chem. Soc.* **125**, 8400–8404 (2003).
- Bracken, C., Carr, P. A., Cavanagh, J. & Palmer, A. G. Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: Implications for the entropy of association with DNA. *J. Mol. Biol.* **285**, 2133–2146 (1999).
- Zhang, Q., Sun, X. Y., Watt, E. D. & Al-Hashimi, H. M. Resolving the motional modes that code for RNA adaptation. *Science* **311**, 653–656 (2006).
- Hayward, S. & Lee, R. A. Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J. Mol. Graph. Model.* **21**, 181–183 (2002).

naturejobs

**JOBS OF
THE WEEK**

What attracts researchers to a specific country or institution? The fame of its researchers, its prestige, the cutting-edge nature of its work and facilities, and the amenities in the surrounding area are clearly factors. But salary also plays a part in the decision.

Globalization has begun to melt away the financial, cultural and infrastructure obstacles that used to keep scientists confined to a handful of research hotspots. Yet salaries are still noticeably higher in a select group of countries, according to a report by the European Commission. After adjustment for the cost of living, the average remuneration of Austria, Australia, Israel, Japan, Switzerland, the Netherlands and Luxembourg are all on a par with that of the United States, at €56,000–62,000 (US\$83,000–92,000). In India, the average living cost-adjusted income for a researcher is €45,207, close to the overall average for the European Union, at €41,500. China, however, falls well below other countries, with an average cost-adjusted researcher salary of just €13,760.

Of course there are some caveats. Salaries vary widely between sectors and fields. A November report from the US Commission on Professionals in Science and Technology notes that this past summer, the country's highest average science or engineering salary for bachelor's graduates went to chemical-engineering graduates, who were offered an average of \$59,361. This is a stark contrast with the \$34,953 offered to life-sciences graduates or the \$41,506 for chemistry graduates.

And more complicated still, countries vary with respect to how quickly salaries rise. The study by the European Commission ranked EU countries' researcher salaries at different career stages, and showed that the rankings dip and rise depending on the amount of experience. The United Kingdom, for example, ranks thirteenth in the 0–4 years of experience window, but rises to ninth for those with 5–7 years of experience and to seventh for the 8–10 years window. When contemplating careers, scientists need to consider not just the field, but also the region and the flux in salary over time.

Gene Russo, acting editor Naturejobs

CONTACTS

Acting Editor: Gene Russo

European Head Office, London

The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:

Andy Douglas (4975)
e-mail: a.douglas@nature.com
**Business Development
Manager:**
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com

Natureevents:

Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:

Muriel Lestringuez (4994)

Southwest UK/RoW:

Nils Moeller (4953)

Scandinavia/Spain/Portugal/Italy:

Evelina Rubio-Hakansson (4973)

Northeast UK/Ireland:

Matthew Ward (+44 (0) 20 7014 4059)

North Germany/The Netherlands:

Reya Silao (4970)

South Germany/Austria:

Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:

Stephen Russell
To send materials use London
address above.
Tel: +44 (0) 20 7843 4816

Fax: +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Jasmine Myer

US Head Office, New York

75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:

Ayako Watanabe
Tel: +81-3-3267-8765
e-mail: a.watanabe@natureasia.com

MOVERS

Jose Alonso, director, Sanford Underground Science and Engineering Laboratory, Homestake, South Dakota



2005-07: Consultant, Lawrence Berkeley National Laboratory and CERN, Switzerland

2000-02: Project leader, Lawrence Berkeley National Laboratory, Berkeley, California

1994-2000: Deputy director for accelerators, Spallation Neutron Source, Oak Ridge National Laboratory, Tennessee

When a family friend gave then high-school student Jose Alonso a tour of the particle accelerators at the Lawrence Berkeley National Laboratory (LBNL) in Berkeley, California, Alonso knew his career would be in physics.

Alonso got his PhD in nuclear physics at the Massachusetts Institute of Technology in Cambridge, and began his career as a junior researcher at Yale University working for a professor with strong ties to the LBNL. It wasn't long before he jumped at the chance to return to the LBNL. There, he and his wife Carol were part of the team that discovered the element seaborgium.

Alonso's evolution into an accelerator physicist reached a pivotal point when he began to explore new applications for accelerators. "Ferretting out different uses for these machines and interfacing with new communities has become a passion during my career," he says.

His first high-profile project was the Bevalac, a linear accelerator addition to an existing Bevatron accelerator at the LBNL. The Bevalac accelerated relativistic heavy ions used in both nuclear science and radiation therapy. Once newer machines made the Bevalac obsolete in 1993, Alonso turned his attention to the materials-science community's desire for an accelerator-based neutron source. The subsequent Spallation Neutron Source at Oak Ridge National Laboratory in Tennessee combined resources and collaborators spread out across six US national labs, and Alonso became its coordinator.

Alonso officially retired from the LBNL in 2002, but continued to take part in projects including the Large Hadron Collider at CERN near Geneva, Switzerland. Now that the Homestake Mine in Lead, South Dakota, is to become a deep-underground science and engineering laboratory, Alonso is heading the effort to reopen the mine and set up the first experiments.

"It's almost as if my career has been leading up to this," says Alonso. His greatest challenge may be managing the expectations of the various scientists. Particle physicists need the overlying rock to act as a shield while they search for rare neutron decay events, but geochemists want to probe it for chemical clues of life's origin. "Jose strives for consensus approaches to grow a broad base of support," says Kem Robinson, LBNL engineering-division director and long-time mentor. Communication, Alonso agrees, is the hardest part of large-scale physics experiments — a skill he has honed throughout his career. ■

Virginia Gewin

BRICKS & MORTAR

How to handle drug withdrawal

Pharmaceutical company Pfizer left a big gap when it decided to move out of Ann Arbor, Michigan, with the loss of hundreds of jobs. But the town responded by setting up a science incubator and planning to diversify.

Local economic development organization SPARK and the University of Michigan, Ann Arbor — with \$1 million from the state — have assumed Pfizer's lease on 3,150 square metres of lab space and are turning it into a science and technology incubator. SPARK is renting a third of the space to biotech firms, with Oncolmmune of Ohio, SensiGen of Ann Arbor and German company Genomatix Software as the first three tenants. University of Michigan researchers working on spin-off companies will occupy the remainder of the space.

"We are not here crying into our beer because of this," says Mike Finney, president and chief executive of SPARK. Instead, he wants to use Pfizer's move out of Michigan — caused partly by restructuring after buying Pharmacia — as a way to diversify the state's economy. Rather than rely on biopharmaceuticals, SPARK has courted high-tech companies in other sectors. Google will eventually hire 1,000 people for its Ann Arbor area operation. Spanish aerospace company Aeronova plans

to hire 600 engineers at an Ann Arbor site. And Toyota and Hyundai are developing technical centres.

Another 186,000 square metres of former Pfizer lab space will be empty by next August, and another 2,000 jobs lost. According to SPARK officials, some are relocating to other facilities, some are retiring, and about 600 hope to stay and find work.

The loss of Pfizer and the shrinking of the Detroit automobile industry have been hard for Michigan, which has the highest unemployment in the United States. Finding tenants to fill 186,000 square metres will be a big challenge. "That's a lot of space," says Steve Forrest, University of Michigan vice-president for research.

SPARK is working with 19 groups of Pfizer employees who want to launch their own companies, and Michigan has put up \$8.5 million in start-up funds. Forrest says that some University of Michigan start-ups have hired former Pfizer employees and expects that more companies will emerge once Pfizer has left. The incubator will allow university companies and private biotechs to interact and share resources.

"We're very strongly in favour of not just funding our own companies, but building up infrastructure for others," Forrest says. ■

Paul Smaglik

POSTDOC JOURNAL

A conference too soon

I enjoy scientific conferences. I can find out what everyone else in my field is up to, and receive advance notice of any exciting new developments. I get a valuable opportunity to discuss technical issues and talk through problems with the real experts. As the sole palaeomagnetist on the staff here in Johannesburg, that's not something I get to do very often. Most importantly for a young researcher like me, I can improve my visibility in the scientific community: future collaborations and jobs will probably come much more easily if people can put a face to the name on an application or proposal.

So when I got an abstract accepted for this year's American Geophysical Union meeting in San Francisco, I was very excited — until reality intervened. It is hardly a short hop from South Africa to California, and my attendance would far exceed my travel budget for this year. As I can afford to go to only one or two big conferences during my postdoc, I'll be spending my money much more wisely, and I'll make much more of an impact, if I've got an exciting story to present on the basis of my work here in South Africa. It's a little early yet to tell that story. So I regret that San Francisco will have to wait. Perhaps next year I'll be ready. ■

Chris Rowan is a postdoctoral student in the geology department at the University of Johannesburg, South Africa.

Life, abundant and with simple joy

Found in translation.

Sarah K. Castle

Eshe Mintz was one of the lucky millions who experienced the martian first contact as it occurred. She was up giving the Batemans' baby, Madeline, a bottle at 09:13 GMT on 7 November 2018.

She'd been following one of the consensus-driven decapodes on the I-TV in the nursery. The decas were her favourite of all the robots dropped on Mars for the mission dubbed Robot Rain. A big, six-legged mother would go as far as it could, then reel out the decas. At half-a-metre long and ten centimetres wide, the ten-legged wonders could go anywhere.

This deca was exploring a lava tube near Tharsis Montes. The views were mostly of rough, black rock, but there was an action vote about every half-hour, which made it more exciting than the longer-ranging surface robots.

The viewers were considering whether to wiggle through a rock pile, hoping to find a side passage, or to continue down the main tunnel. The consensus-tracking bar showed opinions almost equally divided and bouncing towards the side exploration. Everybody loved mapping new passages, but the chatterbox showed strong opinions that they'd fallen too far behind the International Space Cooperative's scientist-driven decas. They'd gone farther down the main tunnel.

Because the side passage was slightly favoured, the deca poked its head into openings in the rock pile while the debate continued. During one of these pokes, Eshe thought she saw something light-coloured, unusual in a basalt tunnel. She leaned forward, tapping the screen to change her vote to favour the side passage, and her torso pushed Madeline's bottle aside, breaking the baby's suck.

Madeline cried. For her first few weeks as the Batemans' nanny, Eshe had taken the crying personally, thinking she'd done something wrong. Now she knew it was just a reaction to some discomfort the baby felt. As soon as she could make it right, the crying would stop. By the time she got the nipple back in, the chatterbox was filled with shouting.

"IT'S ALIVE! IT'S ALIVE!"

The consensus-tracker showed 96% had changed their votes to the side passage at

the same time Eshe had. The consensus was so strong, the ISC decas had turned in their tracks and were coming back.

Eshe maximized the window, her hand shaking on the mouse. The deca had wriggled through the rock pile and come into a wide new tunnel. A line of slumped, lumpy figures stood in the view, glowing with faint white-orange light. They looked like white-sheet ghosts, sagging under the weight of their bulky heads. The deca triangulated to report a height of 1.3 metres.



Still as statues, they might have been stocky stalagmites except for the roughly oval areas on each figure's crown, fading from pale pink to yellow. The colours streamed across their heads, travelling down the tunnel and out of sight, like lights along a runway.

Traffic jumped in the chatterbox. It switched to sample mode, randomly displaying from the submitted comments. Everyone shouted.

"STROMATOLITES?!?!"

"IS IT MOVING?"

"IT'S TRYING TO COMMUNICATE!"

The ISC decas arrived. Their cameras swept the scene methodically. One skittered towards a white figure and poked it with a tiny, tubular sampler. The lights on the figures' crowns quickly faded to a uniform deep red. The comments stopped streaming for long seconds until the colours returned to their paler shades, fading and travelling from figure to figure.

A new comment finally flashed on. "IT SPOKE. IT'S ANGRY."

Links appeared in the chatterbox, leading to sub-discussions with titles like 'What Is It Saying?', 'How Does It Eat?' and 'Can It Know God?'. Madeline fell asleep, but Eshe couldn't take her eyes from the screen. She kept the baby cradled in her crossed legs. Comfortable, Madeline slept, her quick, wet breaths keeping a steady rhythm.

The results streamed in all night. The figures stood on a mass of water ice. The tunnel's atmosphere had a significant methane component. A steady breeze blew through, bringing dust particles from the surface, which caught on the figures' sticky skins. The sampled cells were prokaryote, but with a number of exotic cell types of unknown function.

When the sky turned from black to deep blue outside the nursery window, Eshe was tracking the 'What Is It Saying?' discussion. Thousands of amateurs and dozens of professionals were analysing the sequence of colours modulating along the figures' crowns. Many claimed they'd identified patterns. Some experimented with transforming the flashes into sets of integers, trying to parse mathematical or grammatical syntactic units. Eshe thought she saw one or two figures keeping a steady tempo for short periods of time, but otherwise it looked random.

The morning's first birdsong surprised her. The night hours had flown by. Regan, the Batemans' two-year old, began singing softly in the adjacent room. Eshe imagined the small girl, lying in bed, enjoying the sunrise colours lighting up the walls. Snippets of *Frosty the Snowman*, *Did You Ever See a Lassie* and *Are You Sleeping, Brother John?* were strung together in no particular order. Madeline stirred in Eshe's lap. It would be time for another feeding soon. Before she stood to get the bottle, she submitted her first and only comment to the discussion.

"They sing with simple joy," she typed, resisting the urge to shout. She carried Madeline to the window. Eshe looked up at the brightening sky, laughing quietly with giddy wonder.

Sarah K. Castle (www.skcastle.com) is a geologist and science-fiction author in Flagstaff, Arizona. She's been published in *Analog Science Fiction and Fact* and *Helix: A Speculative Fiction Quarterly*.

JACEY